MÉTODO DE CONTAGEM E CLASSIFICAÇÃO DE TRÁFEGO VEICULAR EM RODOVIAS DE SENTIDO DUPLO COM DATASET PERSONALISADO: UMA ABORDAGEM BASEADA EM INTELIGÊNCIA ARTIFICIAL

Kleberson Rocha[1], Alex Jimenez[1],
Bruna Barcelos[1], Frederico Rodrigues[1], Kaio Mesquita[1], Ludmila Ferreira[1]

[1] Imtraff Engenharia e Mobilidade, Brasil, kleberson.rocha@grupoimtraff.com.br

Resumo

A extração de indicadores rodoviários é um processo complexo que em meio à quarta revolução industrial, carece de implantações viáveis em termos de acurácia e processamento computacional. A inteligência artificial deve ter papel fundamental na mudança desse paradigma. O objetivo deste trabalho é desenvolver um método de contagem, classificação e extração de indicadores de tráfego veicular por intermédio de visão computacional para otimizar processos manuais de coleta de dados e compreensão em tempo real da variação desses indicadores. O método está dividido em quatro etapas: (i) Definição e otimização dos algoritmos; (ii) Delimitação do controle de contagem e classificação; (iii) Implantação dos algoritmos; e (iv) Extração dos indicadores e cálculo comparativo da acurácia dos modelos. Os resultados apontam que a fadiga humana influência negativamente na obtenção dos dados, enquanto que o erro da máquina é constante. Todos os modelos apresentaram acurácia superior a 90% para automóveis de passeio e 70% para caminhões.

Palavras Clave: Visão Computacional; Rodovias; Contagem; Classificação; Otimização.

1 Introdução

A obtenção de indicadores rodoviários é um processo complexo, que necessita de novas metodologias não-invasivas e versáteis. Por isso, o uso de inteligência artificial do tipo visão computacional surge como uma excelente solução para atender à necessidade dentro dos parâmetros exigidos, podendo fornecer uma série de indicadores importantes como volume, velocidade e densidade de tráfego [10, 2]. Dessa forma, tais parâmetros podem ser integrados a um centro de controle de tráfego, permitindo que os Sistemas Inteligente de Transportes (ITS)

sejam dotados de uma ampla gama de recursos como previsão de tráfego em tempo real, maior segurança, planejamento urbano, etc. [9].

Essas novas tecnologias permitem uma melhor visão dos inúmeros fenômenos de mobilidade que geram problemáticas para moradores de núcleos urbanos e rurais. Assim, muitas horas são gastas diariamente com deslocamentos, sendo comum que boa parte desse tempo esteja atrelado a longos engarrafamentos. A análise de tráfego para mudança de rotas e sinalizações, alteração de tempos semafóricos e planejamento de novas rodovias são algumas das maneiras de melhorar, ainda que momentaneamente, os problemas de mobilidade enfrentados pelos usuários [13].

Dessa maneira, o uso de visão computacional surge como uma solução eficaz, capaz de coletar múltiplos indicadores simultaneamente, fornecendo informações mais detalhadas e sem dados esparsos [9, 19]. Isso permite o planejamento de soluções mais rápidas, que forneçam intrinsecamente o conhecimento do local e dos dados de trânsito da região, sem necessitar de muita infraestrutura ou mão de obra. O reconhecimento de objetos é uma área bastante explorada na comunidade científica e tecnológica. Portanto este trabalho usará conceitos de deep learning, unindo modelos de Redes Neurais Convolucionais (CNN) a técnicas de previsão de posicionamento para reconhecer, rastrear e contabilizar veículos em filmagem de ruas e rodovias de forma mais otimizada, além de incluir novas categorias não dispostas na literatura [4]. Desse modo, aponta-se a questão motivadora deste trabalho: Baseado nas técnicas existentes de controle de tráfego e na crescente área de redes neurais, como otimizar o processo de obtenção de indicadores de tráfego e como medir a acurácia desse método para viabilizar estudos de tráfego em rodovias?

Dessa forma, será utilizada a versão padrão do algoritmo YOLOv7 e uma versão personalizada para detecção de veículos em linguagem python, seguido de uma aplicação em um estudo de caso para uma rodovia de duplo sentido e três faixas por sentido. Para o desenvolvimento deste tipo de solução, devem ser considerados fatores-chave que permitam uma boa detecção de objetos, tais como variações bruscas na iluminação, oclusão entre veículos e congestionamento, onde em estudos anteriores, obtiveram uma precisão de até 85% de acurácia na classificação, de forma individual para cada tipo de veículo [2, 12].

Outro ponto importante são os poucos avanços quanto a obtenção de velocidade, uma vez que existe um problema relacionado a taxa de frames e a contabilização desse tempo por passagem de cada veículo, além de não se ter uma distância mensurável apenas pelo vídeo [6, 14].

Dessa forma, a problemática central deste trabalho está relacionada à coleta de dados de forma pouco eficiente ou com baixa precisão devido à preocupação apenas com a aplicação dos modelos desconsiderando as diversas peculiaridades que um vídeo pode apresentar. Portanto, o objetivo deste trabalho é desenvolver um método de contagem, classificação e extração de indicadores de tráfego veicular por intermédio de visão computacional que funcione para uma grande gama de vídeos distintos que permita otimizar processos de coleta de dados e compreensão em tempo real do sistema. Dessa forma, será apresentado um estudo de caso com a implantação e as implicações em se utilizar um dataset personalizado em detrimento de modelos default e contagens manuais. Os resultados que serão apresentados demonstram que a precisão da identificação dos veículos é reflexo da quantidade de objetos no grupo de treinamento. O algoritmo proposto obteve uma precisão de 98% e 78% para identificação correta de carros e caminhões, respectivamente, em um dos sentidos.

Além da problemática levantada na introdução, o trabalho está dividido em mais 4 tópicos: (i) Revisão Bibliográfica e Conceitos; (ii) Proposta metodológica para criação do algoritmo e valiação dos resultados; (iii) Análises comparativas dos resultados de cada algoritmo e controle; e (iv) Conlusões finais.

2 Revisão Bibliográfica

A metodologia de medição de indicadores de tráfego por visão computacional depende inicialmente da capacidade da ferramenta de identificar os veículos nos vídeos. Dessa forma, o uso de Inteligência Artificial (IA), permitem identificar rapidamente de modo semelhante aos seres humanos, os elementos exibidos nas imagens [17]. Para isso, o algoritmo precisa reconhecer

o que está sendo exibido em vídeo, identificando o tipo de objeto e delimitando sua área. Esse processo é denominado delimitação dos quadros ou bounding, em que é selecionado na imagem um elemento, delimitado por uma figura geométrica, que é associado ao objeto analisado no vídeo, que, no caso deste trabalho, tratar-se-á dos veículos detectados nas vias estudadas.

Assim, a detecção de objetos é o primeiro passo para algoritmos do tipo [2]. Para isso, os métodos mais recentes, incluindo os utilizados no presente trabalho (porém otimizados), baseiam-se em subtração por múltiplas camadas. Uma vez separados os elementos de imagem, que pode ser um quadro de um vídeo, os mesmos são analisados por um algoritmo baseado em redes neurais. Estes algoritmos fazem todo um processo de decisão baseado em um aprendizado de máquina que se utiliza de um banco de dados contendo inúmeras imagens classificadas como referência. Portanto, essa metodologia é base dos conceitos adotados no algoritmo utilizado para a detecção dos veículos deste artigo. O YOLO (You Only Look Once) é uma rede neural convolucional única, composta de várias camadas e com funcionamento diferente dos demais algoritmos usuais para este tipo de problema. Proposto inicialmente em 2016, o Yolo não realiza o processo de detecção pelo processo de striding e max pooling para chegar às classes de objetos mais prováveis [7]. Além de mais lento, essa abordagem tem um erro maior na hora de delimitar as áreas de interesse, o que pode acarretar numa confusão entre o plano de fundo e os objetos que de fato estavam sendo identificados, tendo em vista que são extraídas diversas regiões de uma mesma imagem, podendo gerar falsos positivos [11]. Ao invés disso, o processo de detecção é otimizado para separar apenas os elementos propostos e comparar com o uso de uma rede neural convolucional otimizada.

2.1 Rastreamento de Objetos

Ao tratar um vídeo como entrada de um algoritmo de detecção, o primeiro passo é transformá-lo em um conjunto de imagens. Para isso, o algoritmo separa um intervalo de vídeo em uma sequência de quadros ou frames, que são enviados para a rede neural, responsável por processar e identificar os objetos presentes em cada frame independentemente, associando-os com os frames anteriores e posteriores. Assim, para o caso de detecção de objetos que se movimentam, tal qual veículos, é possível separar esses elementos do fundo imóvel da gravação, como a pista e outros elementos da paisagem ao redor.

Pode-se dizer que estes métodos compõem uma das primeiras camadas de uma rede neural convolucional, fornecendo a partir de imagens extraídas de quadros de vídeos as entradas para os demais processos lógicos dela. Com isso, não é só possível identificar os objetos desejados, como também monitorá-los ao longo do vídeo, acompanhando-os quadro a quadro até que estes não sejam mais identificáveis. Os principais obstáculos desse processo são: oclusão, mudança de perspectiva e câmeras que se movimentam [2, 3, 12].

2.2 Redes Neurais e Funções de Ativação

Em 1943 *Walter Pitts* e *Warren McCulloch* publicaram o seu artigo *Logical Calculus of the ideas Immanent in Nervous Activity*, em que propuseram um modelo matemático para representar neurônios biológicos. A tarefa de replicar a complexidade do cérebro humano nas máquinas é o principal objetivo do que hoje é conhecido como redes neurais. A primeira aplicação prática foi utilizada tendo seus conceitos estudados e aplicados com um algoritmo conhecido por *perceptron* [16]. O *perceptron* é uma rede de duas camadas que através de ajustes em seus pesos internos é capaz de realizar tarefas simples como separar dois segmentos. Uma desvantagem deste algoritmo

é a impossibilidade de realizar operações básicas como XOR. A Figura 1 apresenta um exemplo de *perceptron*, tendo sua formulação matemática do tipo: $y = \sum_i^m \quad X_i W_i + b$.

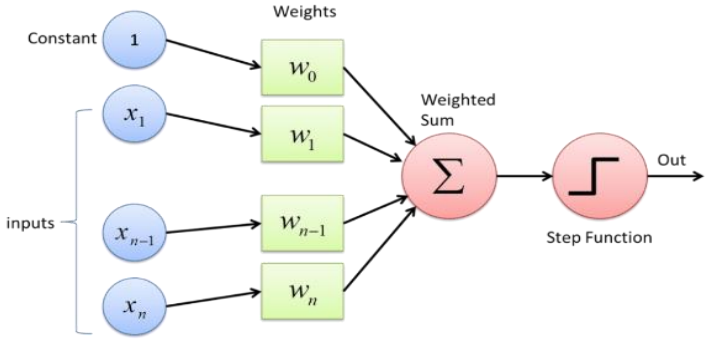

Figura 1 - Ilustração do modelo de um neurônio artificial
Fonte: Adaptado de ROSENBLATT (1958).

Onde a constante b, conhecida como bias, é um peso definido individualmente para cada neurônio da rede. A saída vai ser utilizada ou não, dependendo do seu valor e da função de ativação escolhida. O surgimento do MLP (*Multlayered Perceptron*) foi o que de fato colocou os *perceptrons* no foco de pesquisas de toda a comunidade científica [5]. Neste novo procedimento, a rede era composta de várias camadas que se comunicavam, passando os parâmetros ao longo dos neurônios além de corrigir os pesos utilizados por meio de interações no processo. Ambos os métodos são essenciais para que a rede possa se adaptar às entradas. O treinamento da rede, que é o processo de achar os pesos ideais para cada neurônio, passou a utilizar o que hoje é conhecido como *feedforward* e *backpropagation*. As funções de ativação por outro lado são um paralelo de funcionamento de um neurônio reagindo a uma sinapse. Uma excitação sináptica em um neurônio provocará uma despolarização deste. Se esta despolarização ultrapassar um limite pré-estabelecido, esse neurônio dará continuidade a transmissão dessa sinapse para um próximo neurônio. As funções de ativação são responsáveis por analisar a resposta de cada neurônio de uma rede neural e definir qual será o valor de sua saída. Portanto, as funções de ativação são as responsáveis pela adição de uma etapa não linear, que fará com que a rede seja capaz de aprender problemas de natureza complexa.

2.3 Redes Convolucionais

As redes neurais Convolucionais, também conhecidas como *ConvNets* ou CNN (*Convolutional Neural Networks*), começaram a ganhar protagonismo nos últimos anos, no que tange a paradigmas para visão computacional [1]. O principal motivo da adoção desse modelo de rede é devido a sua capacidade em reconhecer padrões em imagens, como o cérebro humano, e à sua performance. [4] afirma que esse tipo de rede utiliza convoluções no lugar da multiplicação matricial em pelo menos uma de suas camadas. Essa convolução ocorre entre a imagem de entrada e um filtro, ou *kernel*, que é a janela que percorre a imagem fazendo convoluções.

A Figura 2 exemplifica a etapa de convolução entre um *input* e o *kernel*, ou filtro. O input sendo uma imagem preto e branco será composto por vários pixels, sendo representados por 0 e 1, onde 1 representa mais alto brilho e 0 sombra. O *feature map* é o vetor obtido no resultado da operação de convolução entre a entrada e o filtro. Para completar esse valor, o filtro é deslocado ao longo da imagem no processo chamado *striding*, que é definido como um inteiro que define quantos quadrantes serão deslocados para a obtenção do próximo valor a ser preenchido no *feature map*.

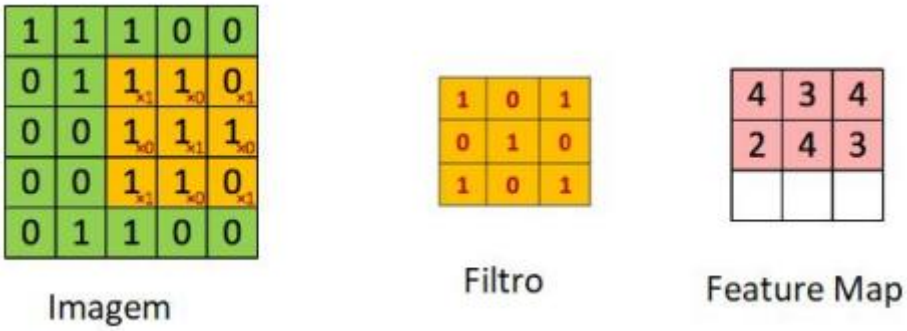

Figura 2 - Processo de convolução entre o input e o filtro.

Depois desse processo é realizado o processo de *pooling*, geralmente o *max pooling*, que subdivide o *feature map* em janelas e seleciona o maior valor de cada janela, descartando os outros valores. Esse processo diminui a complexidade, além de diminuir a chance de *overfitting* [8], ou seja, quando os neurônios passam a decorar o mapeamento e não são capazes de expandir para novos conjuntos de dados.

2.4 YOLO e Tensor Flow

O YOLO (*You Only Look Once*) é uma rede neural composta de várias camadas em pirâmide que são analisadas por uma rede neural convolucional otimizada, para que dessa forma, através do reconhecimento do contexto, seja possível acelerar o processo, uma vez que não apenas o objeto alvo é detectado, mas o plano em que ele está inserido. Essa abordagem permite que o YOLO, que atualmente está em sua sétima versão, obtenha resultados superiores às outras redes. Além disso, por ser de código aberto, a ferramenta permite modificações para treinamentos distintos pela comunidade acadêmica. O processo de detecção do YOLO começa com a subdivisão da imagem em setores. O número de setores é definido previamente e cada um será responsável por detectar um número específico de objetos. As delimitações desses objetos são armazenadas em um *array*, denominado *bounding box*. As *bounding boxes* são caixas delimitadoras que são previstas para cada objeto encontrado na imagem. Esse *array* contém as dimensões da caixa, a posição dela na imagem, a classe de objeto e a probabilidade de certeza da inferência [17]. Para facilitar esse trabalho foram criadas outras bibliotecas e *frameworks* que já possuíam modelos de neurônios e métodos facilitados de implementação. O t*ensorflow* é um desses frameworks com suporte a outras plataformas e a modelos de inferência, desde celulares até supercomputadores com múltiplas GPUs, que possuí facilidade de implementação de redes complexas e simples e suporte a implementação, ou *deploy [7]*.

3 Método da Pesquisa

Na metodologia deste trabalho foi analisado o desempenho e recursos de três ferramentas de visão computacional baseadas nos algoritmos YOLO, na sua quarta e sétima versão, a fim de desenvolver uma solução capaz de realizar contagem, classificação e mensuração de velocidade de veículos a partir de um vídeo base. Vale destacar que a contagem manual também foi aferida para fins de comparação. Baseado neste algoritmo, foram realizadas implementações para otimização da velocidade de processamento e melhoria da acurácia. Uma das modificações foi a substituição do *dataset* (base de dados) padrão por um personalizado com as características da rodovia, de modo a testar a hipótese de que dados de treinamento customizados melhoram a acurácia dos modelos de aprendizado profundo. Uma segunda hipótese testada foi a de que o

fluxo veicular, posicionamento da câmera e sentido da via têm relação com a variabilidade dos resultados.

O método está disposto em seis macro etapas, conforme Figura 3. Em um primeiro momento, foram definidos os principais algoritmos da literatura passíveis de ajuste para melhoramento das predições, seguido da preparação da base de treinamento com a criação de labels utilizando uma plataforma de gerenciamento de datasets (*roboflow*). Foram delimitados cinco tipos de labels (categorias de identificação) para desenho sobre as imagens de treinamento: (i) auto; (ii) moto; (iii) caminhão; (iv) ônibus e (v) rodas.

Dessa forma os dados foram treinados com um conjunto de 484 imagens com mesmo posicionamento de câmera (acima de 3 m no canteiro central) dos vídeos que seriam utilizados para aplicação. O vídeo para aplicação do método foi selecionado com um teste de 1 min de classificação dentre uma amostra de 10 vídeos, sendo selecionado o que obteve maior diferença entre as contagens. Conforme selecionado e aplicado cada um dos modelos, foi realizado o controle manual em velocidade de vídeo 0.5x para garantir a veracidade, além de possibilidade de paralisação. Esse processo foi repetido 3 vezes e considerado como controle a média desses valores [14]. Por fim, foi realizada a comparação com o controle para validação dos dados obtidos, cálculo da acurácia, velocidade média no espaço, fluxo, densidade e classificação.

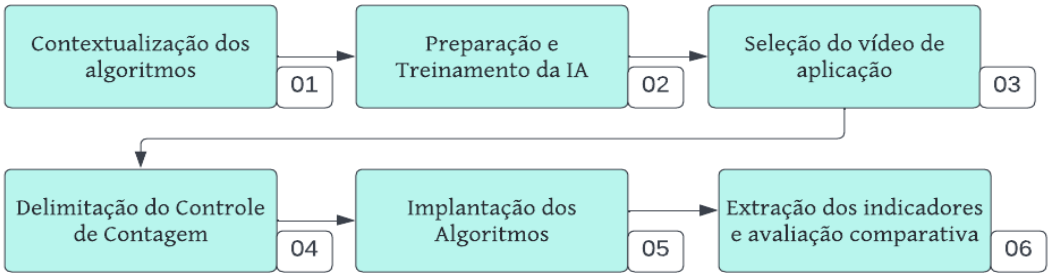

Figura 3 - Metodologia para contagem e classificação veicular otimizada.

Dessa forma, cada uma das aplicações foi realizada extraindo dados de um vídeo de 60 minutos utilizando as metodologias descritas, nos intervalos de 15 min, 30 min e 60 min do vídeo analisado. Vale destacar que a divisão entre veículos leves e pesados, assim como a contagem a cada 15 min, está de acordo com os padrões estabelecidos pelo Departamento Nacional de Infraestrutura de Transportes (DNIT).

O trecho analisado foi gravado no canteiro central de uma rodovia de sentido duplo com três faixas cada. O comparativo foi realizado em uma velocidade normal de reprodução, no qual se considerou a produtividade da máquina em relação ao ser humano, levando em consideração o efeito da fadiga e a perda de concentração, deste último, no decorrer do intervalo proposto. As modificações no algoritmo *default* do YOLO permitiram o desenho de novas faixas de identificação em um mesmo vídeo, contador e classificador em tempo real fixo na tela e extração dos já citados indicadores. O teste de cada algoritmo foi realizado reproduzindo o vídeo uma vez em velocidade normal e exportando os dados através da biblioteca *Pandas* do *python*, mesma linguagem utilizada para estruturação dos algoritmos. De forma paralela, os testes em contagem manual foram feitos por 10 tabuladores, cada um realizando a metodologia de registro das informações de tráfego de contagem e classificação por meio de um contador manual. A seguir são discutidos os pontos principais do algoritmo.

3.1 Detecção de veículos

Os pixels que não variam de intensidade em um espaço temporal pré-definido, comparando-se com frames anteriores, são classificados como integrantes do plano de fundo. Os pixels que

variam significativamente de intensidade, também em relação a frames anteriores, compõem o primeiro plano. O resultado dessa transformação é uma imagem binária composta apenas por dois valores diferentes de pixels, dos quais os que estão em primeiro plano vão representar os objetos que estão se deslocando e os que estão em segundo plano os objetos estáticos, sendo esse o critério adotado para distinguir os veículos dos demais componentes de paisagem presentes na imagem. Para segmentação do vídeo, avançou-se criando um script python para extrair os frames do vídeo na mesma taxa de gravação baseado nos metadados dos arquivos, contendo em média 71,93 quadros por segundo. Outro ponto importante do método proposto é a conversão de todas as imagens para a coloração preto e branco (BW), pois para a identificação de objetos o sistema RGB não influencia, além de ser composto por três matrizes sobrepostas que dificultam a identificação da variação de intensidade dos frames. Portanto, sua conversão para BW reduz a complexidade para duas matrizes sobrepostas que se comportam como uma única matriz variando a intensidade de 0 a 1, em que 0 representa a ausência de cor (preto) e 1 luminosidade máxima (branco).

Posteriormente, foram delimitadas duas linhas de interesse (LI) no vídeo, uma para identificação e contagem primária, e outra para identificação, mensuração de velocidade e contagem secundária. As mesmas seções foram ajustadas pela delimitação dos pixels para distinguir os sentidos de acordo com a ordem de passagem dos veículos, onde um identificador do tipo A-B corresponde ao sentido 1, e identificador do tipo B-A corresponde ao sentido 2, semelhante a um método de contagem com contadores pneumáticos. Os locais das linhas de interesse foram definidos em campo, com uma distância real de seis metros entre elas. É importante destacar que essa delimitação da taxa de frames, o tempo de passagem entre as seções, distância real e distância em pixels das seções são a base para cálculo da velocidade.

Com os elementos da imagem separados, e manipulados, foi necessário introduzir um método que permite relacionar um blob (região de interesse) de um frame, com o do frame posterior, com a finalidade de rastrear o objeto durante a sua passagem pelo vídeo. Para isso, os centróides de cada objeto em movimento são identificados usando os dados gerados pelo YOLO e quando a distância Euclidiana entre o centróide da blob do frame anterior e o centróide da blob do frame atual está abaixo de um determinado valor, é considerado que o objeto do frame anterior é o mesmo do presente. Ou seja, essa propriedade é utilizada na metodologia para determinar os pontos específicos que o corpo do objeto detectado se encontra em relação à imagem analisada, assim como avaliar seu deslocamento ao decorrer do vídeo e, consequentemente, identificar quando o veículo passa pela linha de interesse [5].

3.2 Obtenção e Mensuração dos Indicadores

Uma vez detectados e classificados os veículos, foi realizada a mensuração dos indicadores de tráfego a partir da passagem dos veículos que utilizam a rodovia no vídeo de análise. Essa metodologia foi adotada tanto para a contagem manual como para a contagem por visão computacional. O método, a priori, se baseou em trabalhos que buscaram avançar nos estudos de tráfego através dos avanços da revolução industrial 4.0, sendo a era dos dados e da automatização [3, 12, 15, 2]. Assim, no caso dos experimentos propostos, todo veículo detectado somente será contabilizado ao passar por essa linha que cortará o vídeo horizontalmente, sendo o diferencial desse método, a prevenção da dupla identificação de um veículo permitindo, assim, melhor a acurácia, uma vez que as probabilidades de identificação das duas seções são comparadas. O mesmo critério foi utilizado para a contagem manual, porém os operadores utilizaram botões em uma interface conforme os veículos cruzam as linhas de interesse.

4 Resultados

Os resultados estão subdivididos em três macro tópicos. No primeiro será apresentado os resultados do controle de contagem e classificação veicular seguindo os parâmetros de contagem a cada 15min e por sentido de deslocamento. Ainda neste tópico serão apresentados os resultados das contagens manuais, bem como os erros atrelados a fadiga humana, influenciando na elegibilidade dos resultados encontrados. Por conseguinte, será apresentado os resultados

comparativos dos modelos Yolo4, Yolo7 customizado (Yolo7C) e Yolo7 default (Yolo7D). Por fim, os resultados extraídos serão comparados entre si, validando-os com o controle e analisando o erro de mensuração da contagem e classificação de veículos por métodos de visão computacional em relação à contagem manual. Para seleção do vídeo, conforme apresentado na Seção 3, foi utilizado o de maior discrepância entre a contagem das categorias. Este vídeo obteve uma diferença entre o Yolo7C e Yolo7D, respectivamente, de 62% e 52% para autos (sentido A e B), e, também respectivamente, de 33% e 42% para caminhões (sentido A e B). Coincidentemente foi o vídeo com maior fluxo veicular, corroborando com a hipótese de que o fluxo veicular influência na identificação, sendo necessárias medidas mitigadoras para reduzir os erros nesses casos.

4.1 Contagem Manual e Controle

Para a contagem manual, conforme exposto, foram necessários 10 tabuladores assistindo o vídeo ininterruptamente e salvando os valores de cada categoria em uma planilha *Excel* através de um software para tabulação. A Tabela 1 apresenta o resumo da contagem manual e ao final os resultados do controle de classificação e contagem. Quando os valores agregados por média de todos os tabuladores foram comparados ao controle, o erro máximo de acurácia foi de 21% na contagem de caminhões e o mínimo foi de 2% na contagem de autos, valores aceitáveis, pois na pior das hipóteses 79% e 98% dos valores contados estariam corretos para caminhões e autos, respectivamente.

Tabela 1: Resumo dos resultados da contagem manual e controle

Categoria	Média		Desvio Padrão		Coeficiente de Variação		Controle	
Sentido	A	B	A	B	A	B	A	B
Carro	4141	2956	41	9	1,04%	0,30%	4169	2956
Moto	120	49	2	2	1,67%	4,08%	114	49
Caminhão	103	104	7	9	6,80%	8,65%	129	104

Fonte: Autores.

O problema da prática de contagem manual é que um mesmo vídeo é contado uma única vez por um único tabulador, dessa forma, o erro humano por falta de atenção, percepção e fadiga se acumulam com o passar do tempo. Logo, essa análise deve ser feita comparando cada contagem separadamente ao controle. Quando avaliado separadamente, por tabulador, o erro por contagem em relação ao controle aumenta em 5% para automóveis de passeio e 28% para caminhões, tendo esta última categoria apresentado maior coeficiente de variação em ambos os sentidos. Outro ponto importante é que os menores erros foram referentes ao sentido B (Sul-Norte), no qual os veículos são visualizados pela parte traseira no vídeo e com menor fluxo veicular. Outro fator importante percebido foi que os desvios em relação à média no sentido B apresentaram-se menores do que no sentido A. Embora esses resultados sejam contra intuitivos, tendo em vista que o lado esquerdo do cérebro humano é o responsável pela concentração e raciocínio lógico (SCHERER, 2009), tendo os piores resultados, vale destacar que o fluxo no sentido B (lado direto do vídeo) é 30% inferior que o sentido inverso, ou seja, o impacto do número de elementos e tempo de operação da ferramenta apresentam maiores

efeitos sobre a acurácia dos resultados. Os últimos 15 min apresentaram erros 6% superiores aos outros intervalos de 15min, dando mais evidências sobre o último apontamento.

4.2 Contagem, Classificação e Medição da Velocidade Média através do Yolo

Os algoritmos do Yolo 4 (modelo antigo default), Yolo7D e Yolo7C, originalmente são utilizados para contagem e classificação de autos. Neste trabalho, conforme apresentado na Seção 3, ambos foram automatizados para contagem e classificação de categorias como moto, ônibus e caminhão (além do já citado automóvel). Também foram extraídos indicadores de velocidade, fluxo a cada 15 min e 60 min, e densidade para automóveis e caminhões. A relação fundamental do tráfego foi necessária para verificação deste último indicador.

Tabela 2: Resultados de fluxo veicular a cada 5min – *Yol 4*

FLUXO (Veic./h)	5		10		15		20		25		30		35		40		45		50		55		60		TOTAL	
SENTIDO	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Auto	262	212	275	201	273	203	186	229	195	222	216	207	193	194	227	235	239	215	295	239	303	210	281	218	2945	2585
Moto	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	4
Ônibus	1	1	0	0	0	2	0	0	0	0	3	2	0	3	1	1	0	2	3	0	1	0	0	0	9	11
Caminhão	22	43	14	24	11	26	27	16	23	30	25	24	14	18	17	22	19	26	24	28	31	25	20	17	247	299

Fonte: Autores.

Tabela 3: Resultados de fluxo veicular a cada 5min – *Yolo7C*

FLUXO (Veic./h)	5		10		15		20		25		30		35		40		45		50		55		60		TOTAL	
SENTIDO	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Auto	346	266	357	258	349	249	326	245	319	237	313	223	328	230	368	270	351	260	354	257	357	222	333	232	4101	2949
Moto	10	3	8	3	6	3	11	2	12	6	6	2	5	5	8	4	11	1	12	0	14	4	11	2	114	35
Ônibus	0	0	0	1	0	2	0	1	0	0	0	1	0	0	0	1	0	2	1	0	1	0	0	0	2	8
Caminhão	15	9	11	10	8	12	11	9	14	15	14	12	13	17	13	10	17	9	16	8	16	11	9	8	157	130

Fonte: Autores.

Tabela 4: Resultados de fluxo veicular a cada 5min – *Yolo7D*

FLUXO (Veic./h)	5		10		15		20		25		30		35		40		45		50		55		60		TOTAL	
SENTIDO	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Auto	293	214	286	200	291	196	254	200	229	177	223	175	231	189	294	230	318	216	298	218	286	196	223	202	3226	2413
Moto	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2
Ônibus	1	0	3	0	1	1	1	0	2	2	1	0	3	2	4	1	3	3	6	0	9	1	6	1	40	11
Caminhão	29	34	36	34	21	30	36	32	38	39	44	37	50	33	37	39	30	33	33	33	37	31	31	31	422	406

Fonte: Autores.

Tabela 5: Resultados de velocidade veicular a cada 5min – *Yolo4*

VELOCIDADE (Km/k)	5		10		15		20		25		30		35		40		45		50		55		60		MÉDIA	
SENTIDO	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Auto	44	69	37	77	52	82	76	87	81	82	86	86	38	72	34	74	42	79	69	84	77	86	83	84	60	80
Caminhão	54	61	32	61	60	89	75	84	77	71	87	95	36	53	52	64	56	71	65	74	87	90	91	88	64	75

Fonte: Autores.

Tabela 6: Resultados de velocidade veicular a cada 5min – *Yolo7C*

VELOCIDADE (Km/k)	5		10		15		20		25		30		35		40		45		50		55		60		MÉDIA	
SENTIDO	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Auto	78	78	76	77	76	76	74	78,14	79	75	79	76	79	79	75	76	57	75	67	76	71	77	74	76	74	77
Caminhão	52	70	65	58	58	63	64	60,32	62	64	66	65	65	68	60	57	52	65	66	60	66	64	65	72	62	64

Fonte: Autores.

Tabela 7: Resultados de velocidade veicular a cada 5min – *Yolo7D*

VELOCIDADE (Km/k)	5		10		15		20		25		30		35		40		45		50		55		60		MÉDIA	
SENTIDO	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Auto	62	75	62	74	66	79	53	69	57	71	67	79	44	80	46	76	45	77	48	76	56	78	73	79	57	76
Caminhao	57	68	48	65	63	75	54	61	64	60	71	82	49	58	54	68	52	69	39	68	51	69	44	79	54	68

Fonte: Autores.

As Tabelas 2 a 7 apresentam os indicadores de contagem classificatória de veículos leves e pesados, fluxo horário (somatório das contagens a cada 5 min), velocidade média e a cada 5 min e 60 min baseada nas duas seções espaçadas em 6m. A Figura 5 apresenta o algoritmo de contagem do *Yolo7C* identificando os veículos por sentido e a categoria seguindo a metodologia estabelecida.

O tempo de execução do Yolo4 foi 71% e 27% superior ao do Yolo7C e Yolo7D, respectivamente, tornando-o menos produtivo. Apesar do alto tempo de execução do Yolo 4, o *trade-off* de acurácia e tempo de execução ainda é válido, já que não há necessidade de ter uma pessoa observando a contagem, podendo dar enfoque a outras atividades enquanto o algoritmo é executado. Quanto aos indicadores, todos os algoritmos apresentaram melhores resultados no sentido B. Os modelos Yolo4 e Yolo7D apresentam deficiência em identificar motos e ônibus, porém pode ser contornado utilizando um novo rearranjo da rede convolucional, de modo que as camadas internas apresentam funções e bias que minimizem o erro de previsão, podendo ser melhorada através de funções regularizadas. Os ônibus foram retirados da apresentação final, pois os algoritmos apresentaram 80% de falsos positivos para essa categoria, confundindo-os com vans e caminhões de até 3 eixos, sendo um ponto que necessita de análise mais aprofundada por falta de referências na literatura.

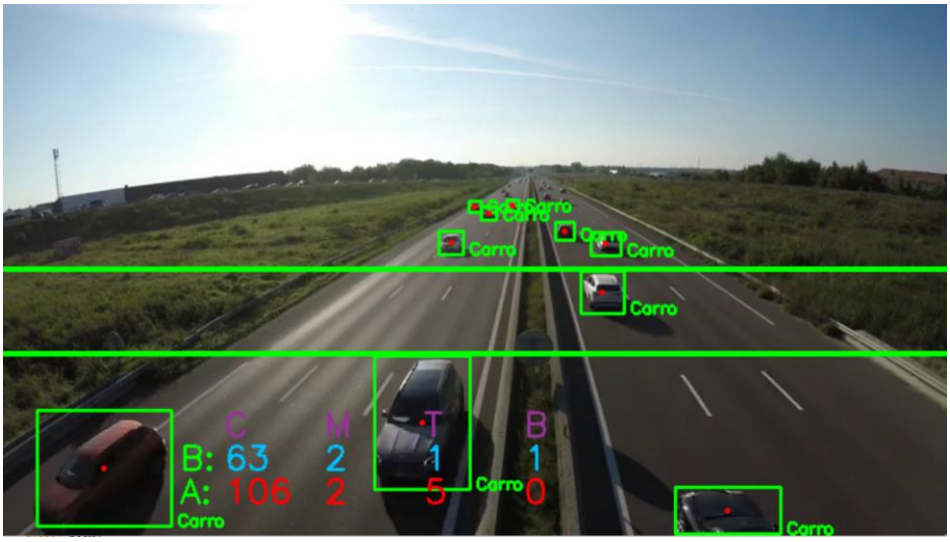

Figura 5: Contagem classificatória – Yolo 7C

4.3 Análise comparativa entre os Métodos

Analisando-se os algoritmos em comparação ao controle (Tabela 8) é possível identificar uma tendência de que os algoritmos sem um conjunto personalizado de treinamento, apresentam maiores erros de previsão. O Yolo7 apresenta uma arquitetura de *Darknet*, em que não detecta um objeto, mas identifica vários elementos ao mesmo tempo em uma mesma imagem, realocando-os em *frames* consecutivos e identificando-os com a mesma classificação. O que pode ser um fator de melhoramento da previsão, não se traduz nos resultados quando se compara os resultados do Yolo4 com o Yolo7D, obtendo taxas de erro para autos acima de 22% e para caminhões acima de 90%. Mas é necessário levar em consideração que a taxa de falsos positivos do Yolo4 é consideravelmente superior à mesma taxa do Yolo7D, o que, no somatório total, acaba contrabalanceando com os falsos negativos que ambos apresentam e ocorre essa falsa sensação de que o Yolo4 é mais preciso que o Yolo7D. Além da melhoria da arquitetura, deve-se levar em

consideração o treinamento especializado dos dados de entrada e posicionamento e oclusão da câmera. Outro ponto levantado é que devido a aparência homogênea dos autos e, no caso do Yolo7C, maior quantidade dos mesmos nas imagens selecionadas para treinamento, sua detecção se deu bem superior para os caminhões, e praticamente nula para as motos. Dessa forma, avalia-se que o fluxo de veículo de uma rodovia apresenta relação com a acurácia dos modelos.

Tabela 8: Tabela resumo com os resultados comparativos

		SENTIDO A						SENTIDO B					
VEÍCULO	ALGORÍTMO	00:00/ 00:15	00:15/ 00:30	00:30/ 00:45	00:45/ 01:00	TOTAL	Variação Controle	00:00/ 00:15	00:15/ 00:30	00:30/ 00:45	00:45/ 01:00	TOTAL	Variação Controle
Carro	YOLO 4	810	597	659	879	2945	-29%	616	658	644	667	2585	-12%
Carro	Controle	1062	973	1068	1066	4169	-	776	706	752	714	2948	-
Carro	Yolo 7 Default	870	706	843	807	3226	-23%	610	552	635	616	2413	-18%
Carro	Yolo 7 Custom	1052	958	1047	1044	4101	-2%	773	705	760	711	2949	0%
Moto	YOLO 4	0	0	0	0	0	-100%	2	1	1	0	4	-91%
Moto	Controle	25	29	21	39	114	-	12	16	12	7	47	-
Moto	Yolo 7 Default	0	0	0	0	0	-100%	0	1	10	0	11	-77%
Moto	Yolo 7 Custom	24	29	24	37	114	0%	9	10	6	6	31	-34%
Caminhao	YOLO 4	47	75	50	75	247	91%	93	70	66	70	299	172%
Caminhao	Controle	34	31	36	28	129	-	27	29	31	23	110	-
Caminhao	Yolo 7 Default	86	118	117	101	422	227%	98	108	105	95	406	269%
Caminhão	Yolo 7 Custom	34	39	43	41	157	22%	31	36	36	27	130	18%

Fonte: Autores.

Os resultados ainda revelam que a taxa de precisão é reflexo direto da quantidade de veículos presentes na preparação do dataset visto que se tinham apenas 249 caminhões, 117 motos e 28 ônibus em oposição aos 2720 carros encontrados nas 484 fotos selecionadas. Por fim, a variação da acurácia em relação ao controle da rodovia apresentou índice de contagem e classificação correta no Yolo7C de 98% e 78% para autos e caminhões no sentido A e 99% e 82% para autos e caminhões no sentido B.

5 Considerações Finais

Diante do exposto, a utilização de inteligência artificial pode auxiliar no melhoramento da acurácia e otimização de processos repetitivos que causam fadiga comumente nos seres humanos. A interação de novas arquiteturas além da CNN deve ser testada, e o treinamento de novos posicionamentos (ângulo e altura) das câmeras, evitando oclusões e ruídos, devem ser implementadas em trabalhos futuros.

Trabalhos comparando vídeos em posicionamentos diferentes para medição de acurácia ou de identificação de objetos com a câmera em movimento devem ser testados. Os resultados apontaram que o fluxo e a diversidade veicular influência na percepção humana e que em alto contingente veicular pode influenciar de forma negativa na acurácia desses resultados. A detecção de veículos como automóveis e caminhões se mostra mais eficiente quando realizada pela parte traseira deles e no sentido direito da tela. Além disso a detecção de rodas pode auxiliar na classificação de veículos pesados por número de eixos.

As análises com modelo personalizado obtiveram melhores resultados em todas as categorias, validando as hipóteses levantadas. O ganho em processamento, cálculo de indicadores em tempo real e a não necessidade de acompanhamento da extração dos valores, tornam os algoritmos de inteligência artificial viáveis de serem utilizados em larga escala, tanto para uso em dados massivos quanto para pesquisas de tráfego veicular. Portanto esse trabalho, buscou evidenciar e dar contribuição em relação a literatura existente com um método de controle de qualidade dos dados e implementação de modelos personalizados para cada tipo de classificação, bem como a obtenção automatizada dos indicadores necessários aos estudos de tráfego.

Referencias:

[1] ANDERSON, Michael; IANDOLA, Forrest; KEUTZER, Kurt. *Quantifying the energy efficiency of object recognition and optical flow*. California Univ Berkeley Dept Of Electrical Engineering And Computer Sciences, 2014.

[2] AZIMJONOV, J.; OZMEN A. *Vision based vehicle tracking on highway traffic using bounding box features to extract statistical information*. Computer and Eletrical Engineering, 2022.

[3] BARCELLOS, P. *A novel video based system for detecting and counting vehicles at user-defined virtual loops*. Expert Systems with Applications, v. 42, n. 4, p. 1845–1856, 2015

[4]BENGIO, Yoshua; GOODFELLOW, Ian; COURVILLE, Aaron. *Deep learning*. Massachusetts, USA:: MIT press, 2017.

[5] CASTRO, L. N. *Análise e síntese de estratégias de aprendizado para redes neurais artificiais*. Campinas: FEEC, UNICAMP. Dissertação de Mestrado-Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, p. 250, 1998. p. 12, 14, 16, 17

[6] CHANG, F.; CHEN, C.-J.; LU, C.-J. *Computer vision and image understanding*. Freedom Collection Journals, p. 206–220, 2003.

[7] GE, L.; DAN, D.; LI, H. *An accurate and robust monitoring method of full-bridge traffic load distribution based on YOLO-v3 machine vision*. Structural Control and Health Monitoring, v. 27, n. 12, 1 dez. 2020.

[8] HAWKINS, Douglas M. *The problem of overfitting. Journal of chemical information and computer sciences*, v. 44, n. 1, p. 1-12, 2004.

[9] HUANG QIUYANG. *Citywide road-network traffic monitoring using large-scale mobile*. Neurocomputing, p. 135–146, 2021.

[10] LI, S. *et al. Vehicle counting and traffic flow parameter estimation for dense traffic scenes*. *IET Intelligent Transport Systems*, v. 14, n. 12, p. 1517–1523, 1 dez. 2020.

[11] LIN, T.-Y. et al. *Feature Pyramid Networks for Object Detection*. 9 dez. 2016. Disponível em: <http://arxiv.org/abs/1612.03144>.

[12] LIU, F.; ZENG, Z.; JIANG, R. *A video-based real-time adaptive vehicle-counting system for urban roads*. PLoS ONE, v. 12, n. 11, 1 nov. 2017.

[13] MOREIRA, M. E. P.; NETO, W. A. P. *Proposição de um método de levantamento e análise de dados para um diagnóstico de um corredor viário urbano*. PLURIS 2006.

[14] RODRÍGUEZ, R. *Analysis of Statistical and Artificial Intelligence Algorithms for Real-Time Speed Estimation Based on Vehicle Detection with YOLO*. Applied Sciences (Switzerland), v. 12, n. 6, 1 mar. 2022.

[15] ROH, S.; SHIN, J.; SOHN, K. *Measuring traffic volumes using an autoencoder with no need to tag images with labels*. Electronics (Switzerland), v. 9, n. 5, 1 maio 2020.

[16] ROSENBLATT, F. *The perceptron: a probabilistic model for information storage and organization in the brain*. Psychological review, v. 65, n. 6, p. 386, 1958.

[17] REDMON, J. *You Only Look Once Unified Real Time Object Detection*. University of Washington, 2016.

[18] SCHERER, L. C. (2009). *Como os hemisférios cerebrais processam o discurso: evidências de estudos comportamentais e de neuroimagem*. Linguagem e cognição: relações interdisciplinares. Porto Alegre: EDIPUCRS, 77-102.

[19] UNZUETA, L.; NIETO, M. *Adaptative Multicue Background Subtraction for Robust Vehicle Couting and Classification*. IEE, p. 527–540, 2012.