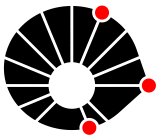

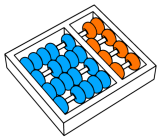

Detecção e Classificação de Veículos com Múltiplos Sensores

Artur Abreu Hendler *Lucas Wanner*

Relatório Técnico - IC-PFG-23-64
Projeto Final de Graduação
2023 - Dezembro

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

Detecção e Classificação de Veículos com Multiplos Sensores

Artur Hendler*

Resumo

Os sistemas tradicionais de monitoramento de trânsito, baseados predominantemente em modelos visuais, podem possuir um alto custo de banda de transmissão e processamento, além de não operarem tão bem durante a noite por conta de uma iluminação reduzida. Este trabalho aborda essa limitação propondo uma abordagem de detecção de veículos baseada na integração de sensores, especificamente microfones e acelerômetros. Foi possível coletar dados reais de uma via e, após tratamentos e extração de features, treinar modelos de aprendizado de máquina capazes de detectar e classificar veículos a partir de suas assinaturas sonoras. A análise dos resultados obtidos aqui busca oferecer informações úteis para aprimorar os sistemas de monitoramento de trânsito. Limitações, incluindo a dependência do cenário específico, condições meteorológicas e diversidade de veículos, indicam oportunidades para futuras melhorias visando uma aplicação mais robusta em diversos ambientes rodoviários.

1 Introdução

Neste trabalho foi explorada a detecção e classificação de veículos utilizando os sensores acelerômetro, microfone e câmera. Com diversos *setups*, a capacidade de cada um deles foi explorada e, no final, foram utilizados para capturar uma grande quantidade de dados e combinados para treinar um modelo de *machine learning* capaz de classificar e detectar alguns tipos de veículos apenas com seu áudio.

1.1 Contextualização do problema

Os sistemas de segurança e monitoramento de trânsito desempenham um papel crucial na garantia da segurança pública e na gestão eficiente do fluxo veicular. Tradicionalmente, a detecção de veículos nesses sistemas tem sido predominantemente baseada em modelos visuais, utilizando câmeras para capturar informações visuais do ambiente rodoviário. No entanto, essa abordagem enfrenta desafios significativos, especialmente quando confrontada com condições noturnas [7].

Durante a noite, a eficácia dos modelos baseados exclusivamente em imagens é notavelmente reduzida devido à limitada visibilidade e variações nas condições de iluminação. Este cenário cria uma lacuna na capacidade de detecção, comprometendo a eficiência global dos sistemas de monitoramento. Além disso, o processamento de grandes volumes de dados visuais pode ser computacionalmente custoso e também, oneroso para a rede, dado um alto tráfego de dados de imagens em sistemas onde a detecção é distribuída [5], tornando esses modelos menos acessíveis e eficazes em cenários práticos.

Assim, surge a necessidade de explorar abordagens alternativas e complementares. Este trabalho propõe uma modo de detecção baseada na integração de sensores multimodais, especificamente utilizando microfones, frequentemente incorporados em câmeras de segurança, e acelerômetros. A

*Inst. de Computação, UNICAMP, 13083-852 Campinas, SP. a231713@dac.unicamp.br

combinação desses sensores visa superar as limitações inerentes aos modelos visuais tradicionais, proporcionando um método mais eficaz e economicamente viável para a detecção e classificação de veículos, mesmo em condições desafiadoras como a escuridão noturna. Ao comparar os resultados obtidos por meio desses sensores com os modelos baseados em câmeras, busca-se oferecer informações úteis para o aprimoramento dos sistemas de monitoramento de trânsito.

1.2 Objetivos do trabalho

A ideia central deste trabalho foi explorar o comportamento dos sensores na detecção de veículos e, posteriormente, a classificação deles. Uma vez confirmada a viabilidade dos sensores, o interesse adicional era integrá-los para uma detecção conjunta.

A estratégia para atingir esse objetivo foi dividida em etapas. Inicialmente, foi estabelecido um *setup* experimental para avaliar a viabilidade dos sensores em um ambiente controlado, isto é, em uma mesa, com carros em miniatura e com algum isolamento acústico. Esse teste preliminar visou entender as capacidades e limitações desses sensores.

Posteriormente, um *setup* mais definitivo foi desenvolvido para coletar dados reais das ruas, buscando observar quais complicações extras poderiam surgir e como era a variabilidade dos dados gerados pelos veículos. Além disso, comparou-se a performance dos sensores com aquela observada inicialmente.

Para gerenciar a grande quantidade de dados gerada na etapa do *setup* definitivo, foi montada uma *pipeline* de processamento. Essa *pipeline* incluiu desde a detecção inicial de veículos pelas imagens até a extração de características relevantes para alimentar modelos de *machine learning*.

Dessa forma, o objetivo desse trabalho foi propor uma abordagem alternativa para a detecção de veículos, ou seja, comparar os resultados obtidos aqui com os modelos baseados unicamente em câmeras. Durante o trabalho, em paralelo à montagem do *setup*, também foi realizada uma revisão bibliográfica para orientar quanto às dificuldades dessas abordagens.

2 Revisão Bibliográfica

A seção de Revisão Bibliográfica apresenta uma análise abrangente sobre os sensores empregados na detecção de veículos, abordando categorias como Detectores de Tráfego *In situ*, Redes Sensoriais de Veículos (VSNs) e Processamento de Imagens e Vídeos. Além disso, são explorados estudos específicos relacionados à detecção em condições adversas, destacando abordagens inovadoras, como a proposta por Bo-Jhen Huang para a Hsuehshan Tunnel em Taiwan e a investigação de Shiferaw H. sobre a detecção de vibrações do solo por meio de acelerômetros de *smartphones*. Essa seção oferece uma visão ampla do estado atual das tecnologias e pesquisas relacionadas à detecção de veículos.

2.1 Sensores para detecção de veículos

A escolha de um sensor é uma das etapas mais importantes quando se desenha um sistema de monitoramento de veículos, há uma grande gama de sensores disponíveis para esse propósito, como apresentado por Jain, N. K. et al em [7]. Nesta artigo, os pesquisadores dividem os sensores em três categorias macro: Detectores de tráfego *In situ* (sensores montados nas vias), Redes sensoriais de veículos (táxis e ônibus que compartilham informações) e Processamento de imagens e vídeos.

Na primeira categoria, Detectores de Tráfego *In situ*, encontramos sensores que podem ser subdivididos em duas tecnologias distintas: Intrusiva e Não Intrusiva.

Intrusiva: Esses detectores são fisicamente instalados na ou abaixo da superfície da estrada, o que, por vezes, pode causar perturbação ao tráfego. Exemplos incluem magnetômetros embutidos, detectores de tubos pneumáticos, laços indutivos e sistemas *Weigh-in-Motion* (WIM).

Não Intrusiva: Sensores desta categoria são montados acima da superfície da estrada, minimizando a perturbação ao tráfego. Eles abrangem métodos manuais (agentes de trânsito), coleta de dados por vídeo, detecção infravermelha passiva ou ativa, radares de micro-ondas, detectores ultrassônicos, detectores acústicos passivos, detectores a laser e fotografia aérea.

Em seguida, na categoria Redes Sensoriais de Veículos (*Vehicular Sensor Networks* - VSNs), dispositivos de sensoriamento móvel são acoplados à automóveis como táxis e ônibus. Esses dispositivos, conectados entre si e ao centro de monitoramento de tráfego, utilizam comunicações sem fio veículo-a-veículo ou veículo-infraestrutura para transmitir dados. Os dispositivos em automóveis são frequentemente referidos como Sistemas de Localização Automática de Veículos (AVL), que podem fornecer informações pontuais ou contínuas sobre a posição deles e, após análises, sobre as condições de trânsito de modo geral.

Por fim, a categoria Processamento de Imagens e Vídeos destaca a ampla aplicação de técnicas de monitoramento de tráfego usando câmeras. O processamento de vídeo ao vivo em interseções, por exemplo, possibilita a estimativa de densidade de tráfego e classificação de veículos [11]. Essas informações são essenciais para a gestão eficiente do tráfego, sincronização de semáforos e, consequentemente, a redução de congestionamentos e acidentes.

Praticamente, embora detectores intrusivos sejam valiosos em aplicações de tráfego devido à sua resposta rápida, baixo custo de operação e manutenção, eles apresentam desvantagens, como a necessidade de escavação nas estradas, custos mais elevados de instalação e informações limitadas em comparação a outros métodos de monitoramento de tráfego [10].

2.2 Detecção e classificação de veículos em condições adversas

Nesta seção, examinamos trabalhos relevantes na detecção e classificação de veículos, com foco especial naqueles que trataram cenários de baixa luminosidade ou que fizeram uso de acelerômetros.

2.2.1 Análise do Tráfego no Túnel de Hsuehshan, por Bo-Jhen Huang

Bo-Jhen Huang abordou a complexidade da detecção de veículos em condições desafiadoras [6], utilizando a Hsuehshan Tunnel em Taiwan como cenário. Este túnel, inaugurado em 2006, é o mais longo de Taiwan e foi projetado para mitigar congestionamentos no tráfego [17]. No entanto, a segurança rodoviária neste túnel extenso é uma preocupação significativa graças a sua reduzida iluminação, já sendo alvo de outros estudos[18].

O trabalho destaca que, atualmente, várias câmeras monitoram o túnel, mas a eficácia desses sistemas é afetada pela qualidade das câmeras e pela iluminação no ambiente. O monitoramento manual torna-se desafiador, especialmente durante condições de congestionamento, levando à necessidade de um sistema de monitoramento de tráfego inteligente.

Huang propôs um sistema de detecção baseado em *background subtraction* e na aplicação de uma *Deep Belief Network* (DBN). A metodologia incluiu a coleta de amostras positivas de veículos, utilizando métodos de subtração de fundo e avaliação de diferentes algoritmos para construir modelos eficazes. A normalização de imagens foi realizada para reduzir a complexidade computacional, convertendo imagens coloridas em tons de cinza, aplicando equalização de histograma e *thresholding*.

Os resultados do estudo indicaram uma taxa de acerto notável de 96.59%, demonstrando a eficácia do método proposto. Huang enfatizou a importância de coletar mais dados de treinamento

para aprimorar ainda mais o modelo, visando uma detecção em tempo real no futuro.

2.2.2 Estudo das vibrações causada pelo tráfego, por Shiferaw H.

O trabalho de Shiferaw [13] direciona sua atenção para a preocupação associada às vibrações do solo induzidas pelo tráfego, particularmente em áreas urbanas com estradas danificadas, estreitas e proximidade residencial. O potencial de causar danos variados a edifícios [8], desde pequenas rachaduras até falhas estruturais, torna essas vibrações uma área de pesquisa de interesse.

O autor propõe uma abordagem mais prática ao empregar sensores de *smartphones*, notadamente os acelerômetros integrados nesses dispositivos, para a aquisição de vibrações externas causadas pelo tráfego. A escolha desses sensores é fundamentada na acessibilidade generalizada de *smartphones*, apresentando-se como uma ferramenta potencialmente valiosa para medições contínuas e monitoramento da saúde estrutural. É necessário, no entanto, destacar importantes limitações em comparação com sensores profissionais de vibração. Diferenças marcantes na faixa de medição e frequência de aquisição. Enquanto a resolução dos celulares varia de 0.1 a 0.0001g, os sensores profissionais de aceleração atingem resoluções de até 10^{-9}g [3].

Shiferaw, em sua avaliação experimental da aquisição de vibração do solo por meio de *smartphones*, testou todas as combinações que eram possíveis dos veículos disponíveis (caminhonete, caminhão e trator com rolo compressor), velocidades (25 e 50km/h), tipos de pavimento (asfalto liso, pedregulhos e asfalto com danos) e distância do acelerômetro (1 e 2 metros). Em sua análise, foi possível detectar vibrações em quase todos os casos, exceto naquele com a caminhonete em uma velocidade mais baixa, ou com o posicionamento do sensor mais distante de seu percurso. Nesses cenários, o autor apontou que a sensibilidade do sensor não permitiu a diferenciação do sinal capturado frente ao ruído inerente ao acelerômetro, deixando claro a limitação do *smartphone*.

3 Metodologia

Nesta seção são discutidas as configurações experimentais usadas para explorar os sensores e, no final, coletar os dados utilizados para treinar modelos de *machine learning*. Adicionalmente, também são expostos alguns detalhes técnicos dos dispositivos utilizados.

3.1 *Setup* Experimental

A etapa inicial do estudo foi conduzida por meio de um experimento controlado, visando avaliar a eficácia do acelerômetro, incorporado em dispositivos móveis, na detecção de vibrações geradas pela passagem de veículos em escala. Este procedimento experimental proporcionou dados para uma análise da resposta do acelerômetro em comparação com dados de áudio e vídeo.

O experimento envolveu o posicionamento de uma mesa na qual carrinhos em escala eram lançados suavemente em proximidade a um dispositivo móvel, neste caso, um celular, figura 1. A escolha da mesa baseou-se em sua baixa rigidez em comparação com superfícies como o asfalto, buscando facilitar a transmissão das vibrações provocadas pelos carrinhos para o dispositivo. O celular foi configurado para gravar vídeo e áudio enquanto registrava simultaneamente as vibrações por meio de seu acelerômetro interno.

O vídeo do celular contou com um áudio amostrado em 44KHz e o acelerômetro, por sua vez, foi capaz de registrar variações de até 0.005 m/s^2 à uma taxa de amostragem de 100Hz. Por se tratar apenas de um teste de hipótese, foram feitos nove apenas sete lançamentos. O áudio coletado pelo vídeo foi então extraído e alinhado temporalmente aos registros do acelerômetro, utilizando para isto, o relógio interno do dispositivo.

Figura 1: *Setup* Inicial do experimento. Mesa, carros modelo a serem lançados e celular para captura

3.2 *Setup* Definitivo

Após a avaliação no ambiente controlado, foi projetada uma nova configuração para monitorar a passagem de veículos em condições mais próximas do tráfego real. Uma câmera de segurança foi fixada em um ponto alto da residência para capturar vídeos da rua. Simultaneamente, um dispositivo móvel com um acelerômetro foi posicionado próximo ao solo, a cerca de 2,5 metros da rua, figura 2.

As filmagens foram extraídas da câmera pela rede local usando o protocolo RTSP para análise inicial. Posteriormente, foi incorporado um Cartão SD para a coleta prolongada de dados de vídeo. Para a captura dos dados do acelerômetro, foi utilizado o aplicativo PhyPhox [16], que disponibiliza os dados em formato CSV, fornecendo informações detalhadas em cada eixo. O acelerômetro utilizado permaneceu o mesmo do caso anterior, com uma sensibilidade de 0.005 m/s^2 e uma taxa de amostragem de 100Hz. A câmera de segurança, por sua vez, proporcionou um vídeo no formato 640x340, gravado a 15 quadros por segundo, com áudio amostrado em 16 kHz.

Como uma etapa de pré-processamento de dados, foi necessário alinhar temporalmente os dados dos sensores, como no último caso. Além disso, no caso do acelerômetro, foi calculada uma aceleração absoluta, combinando os dados de múltiplos eixos e descontada a aceleração constante da gravidade.

Figura 2: *Setup* Definitivo do experimento

4 Estudo de Viabilidade com Acelerômetro

No teste de mesa, como esperado pela proximidade dos veículos com os sensores e pela natureza controlada do experimento, foi trivial observar a passagem dos carrinhos através de ambos os sensores apenas observando a amplitude em função do tempo, figura 3. Além disso, no caso do som, também foi possível fazer uma análise no domínio da frequência, observando a mudança de algumas características como o centroide espectral e o *rollof* espectral para cada trecho amostrado, figura 4.

Para o teste realizado na rua, no entanto, os resultados não foram tão promissores. A amplitude em função do tempo captada pelo acelerômetro para uma amostra de dois minutos pode ser observada na figura 5, nessa figura (15 segundos do total), temos os trechos destacados (apenas para ilustração) que mostram a passagem de veículos. Realizando uma media no valor da amplitude para trechos de 4 segundos no decorrer de toda a amostra capturada (tempo aproximado da passagem de um veículo), obtêm-se 0.011 $\pm$0.004 m/s^2 para trechos sem passagem de veículo e, 0.011 $\pm$0.003

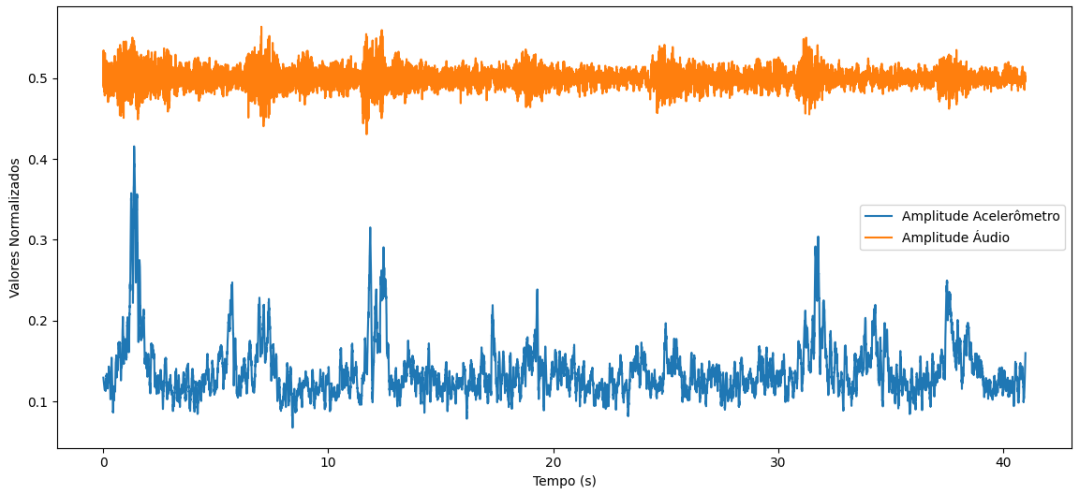

Figura 3: Amplitude normalizada do sinal do áudio e do microfone em função do tempo para o teste de mesa

m/s² e 0.012 ±0.003 m/s² nos momentos de passagem dos veículos destacados.

No caso dos dados capturados pelo microfone interno da câmera, pode-se observar mudanças na amplitude no decorrer da amostra para os mesmos momentos de passagem de veículos (figura 6) e, fazendo uma análise no domínio da frequência (figura 7), é possível ver características mais distintas quando há, por exemplo, a passagem de uma motocicleta (graças ao seu ruído mais agudo), indicando também ser prolífico uma extração de dados nesse domínio.

4.1 Limitações e conclusões sobre a viabilidade do uso do acelerômetro

O estudo de Shiferaw [13] enfatiza o potencial prático dos acelerômetros em *smartphones* para a detecção de vibrações causadas pelo tráfego. Contudo, é necessário reconhecer as limitações intrínsecas desses sensores em comparação com instrumentos profissionais. Como dito anteriormente, a sensibilidade do acelerômetro presente em *smartphones* é entre oito e cinco ordens de grandeza menor quando comparada a dispositivos profissionais.

No estudo de Shiferaw, em algumas situações como a de passagem de veículos menores ou mais distantes do sensor utilizado, limitações de detecção foram encontradas. Essas limitações identificadas pelo autor ressoam com os resultados preliminares deste trabalho, onde a diferenciação entre o ruído do sensor e o sinal gerado pela passagem de veículos se mostrou desafiadora e praticamente inviável no caso de veículos ainda mais leves como motocicletas. Afinal, os valores mensurados para os veículos são quase os mesmos daqueles de quando não há tráfego e com uma grande folga de incerteza.

Concluímos, portanto, que embora os acelerômetros de *smartphones* possam ser ferramentas acessíveis e práticas para a detecção de vibrações em contextos controlados, como o teste de mesa, suas limitações tornam-se mais evidentes em ambientes complexos, como ruas movimentadas. Essas considerações influenciam diretamente a viabilidade do uso do acelerômetro para detecção e classificação de veículos.

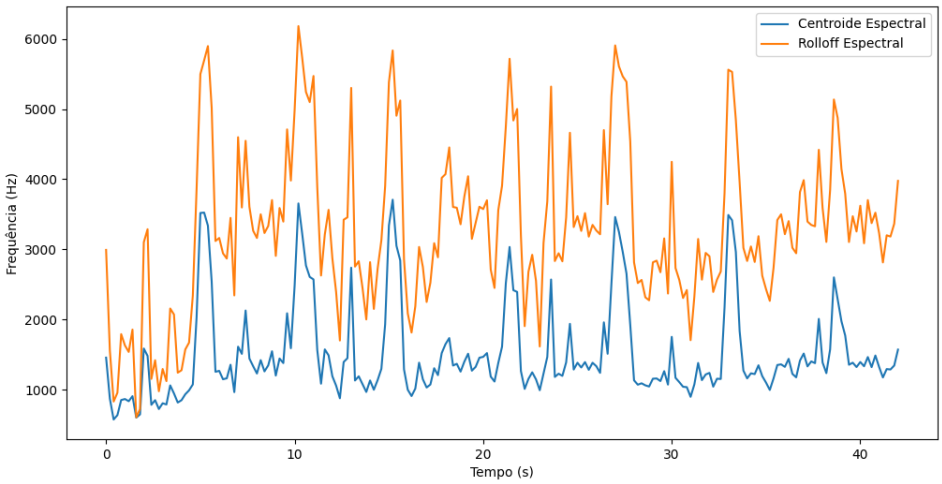

Figura 4: Características no espectro da frequência do sinal para o teste de mesa

5 Processamento de Imagem e Áudio

Dado a dificuldade encontrada no uso do acelerômetro, o foco deste trabalho passou a ser o áudio e o vídeo obtidos através da câmera. A próxima seção descreve como um volume maior de dados foi extraído da câmera visando a separação de trechos de áudio com e sem a passagem de veículos para posterior uso em modelos de *machine learning*.

Com a câmera configurada para capturar vídeos em um cartão de memória, ela foi apontada para a via e deixada por 13 dias realizando esses registros. Por decisões do fabricante, as gravações são armazenadas como trechos de dois minutos, resultando em aproximadamente 9,3 mil vídeos.

5.1 Pipeline de Processamento de Dados

O propósito final do tratamento dos dados consiste em trechos de quatro segundos de áudio que possuam uma *label* indicando a presença ou ausência de um veículo em movimento. Para alcançar tal objetivo, é necessária a detecção automática da passagem de um veículo durante os vídeos e o subsequente salvamento do trecho de áudio correspondente.

Nesse contexto, optou-se pela utilização do modelo de classificação de imagens pré-treinado YOLOv8 [9]. Este modelo, altamente eficaz, inclui entre suas diversas possibilidades de classificação, as classes "car", "motorcycle"e "truck". Desenvolveu-se um programa que processa os vídeos e, a cada segundo (15 *frames*), seleciona um *frame* para determinar a presença ou ausência de um veículo nele.

Após a detecção do veículo, são salvos a imagem (*frame*) responsável pela detecção e 4 segundos de áudio (dois antes e dois depois do momento da detecção). A *timestamp* da imagem e o tipo de veículo são utilizados como nome do arquivo.

A abordagem adotada, após o ajuste de alguns parâmetros como o *threshold* da detecção e a área da imagem utilizada para análise do modelo, demonstrou bons resultados, registrando com precisão a passagem de veículos e apresentando poucos falsos-positivos. Contudo, um desafio encontrado foi a lentidão do processo, considerando a quantidade de dados utilizada. Para otimizar esse aspecto, implementou-se, por meio da biblioteca de Python Subprocess, um paralelismo baseado em *pool de*

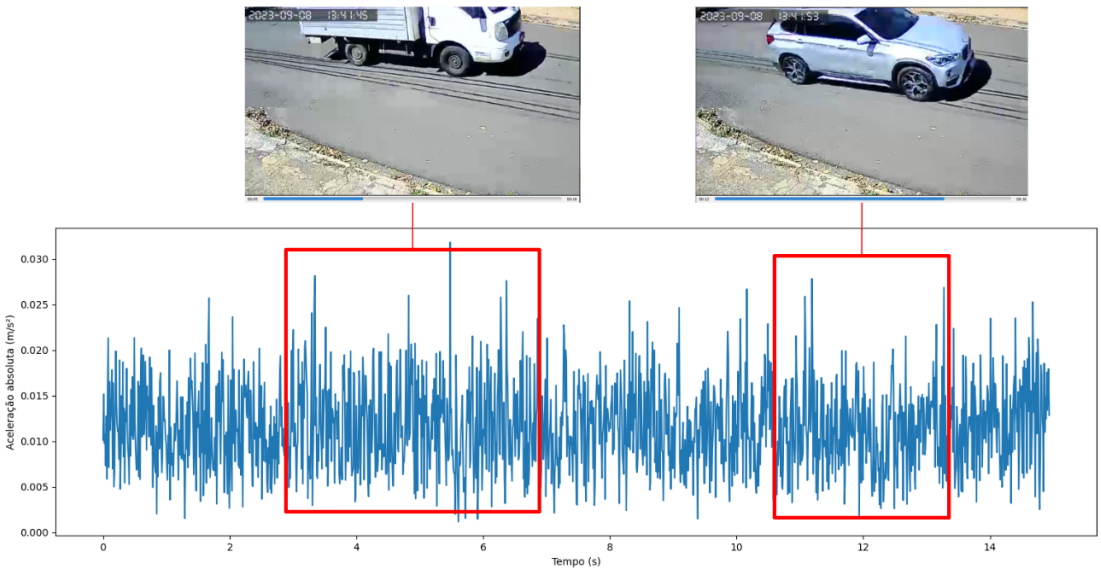

Figura 5: Amplitude do acelerômetro em função do tempo para a amostra de rua

threads, com cada núcleo responsável pela análise individual de cada vídeo.

Adicionalmente, como melhoria de maior impacto, adicionou-se uma camada de detecção de movimento. Nesta camada, antes de realizar a detecção de veículos pelo modelo de YOLOv8, o *frame* atual é comparado com o anterior e é mapeada a área em que houve uma mudança significativa nos pixeis, sendo acionado o modelo apenas no caso positivo. Esta alteração, além de trazer o ganho de desempenho, se mostrou proveitosa para detectar os casos de falso-positivos causados por carros estacionados.

Na figura 8, tem-se um diagrama onde é possível ver de maneira mais clara o caminho dos *frames* de um vídeo na detecção.

5.2 Dados Obtidos

De todos os vídeos coletados, foi utilizado para a geração dos áudios apenas aqueles diurnos, afinal, durante os testes foi observado que a câmera, para melhorar a exposição de sua filmagem, aumenta consideravelmente o tempo do seu obturador (digital), o que causa uma perda muita significativa nos detalhes necessários para a detecção do veículo por parte do modelo YOLOv8. Além disso, a câmera também passa a operar em preto e branco, outro dificultante. Uma comparação entre a passagem diurna e noturna de um carro pode ser vista na figura 9.

Com os vídeos diurnos, foi possível extrair amostras de 3100 carros, 370 motocicletas e 56 caminhões. Além disso, foram também extraídos um número maior que os demais, mas arbitrário de 20.000 *frames* vazios, visando o treinamento do modelo no caso negativo da passagem de um veículo.

5.3 Extração de Features

Com os trechos de áudio coletados é interessante, para o treinamento de modelos baseados nesses dados, extrair as características mais importantes. Para isso, foi utilizada a biblioteca para Python, librosa [12]. Com ela, foram extraídas dos trechos, as features: Potência RMS, desvio padrão da potência RMS, centroide espectral, Desvio padrão do centroide espectral, *zero-crossing rate* (taxa

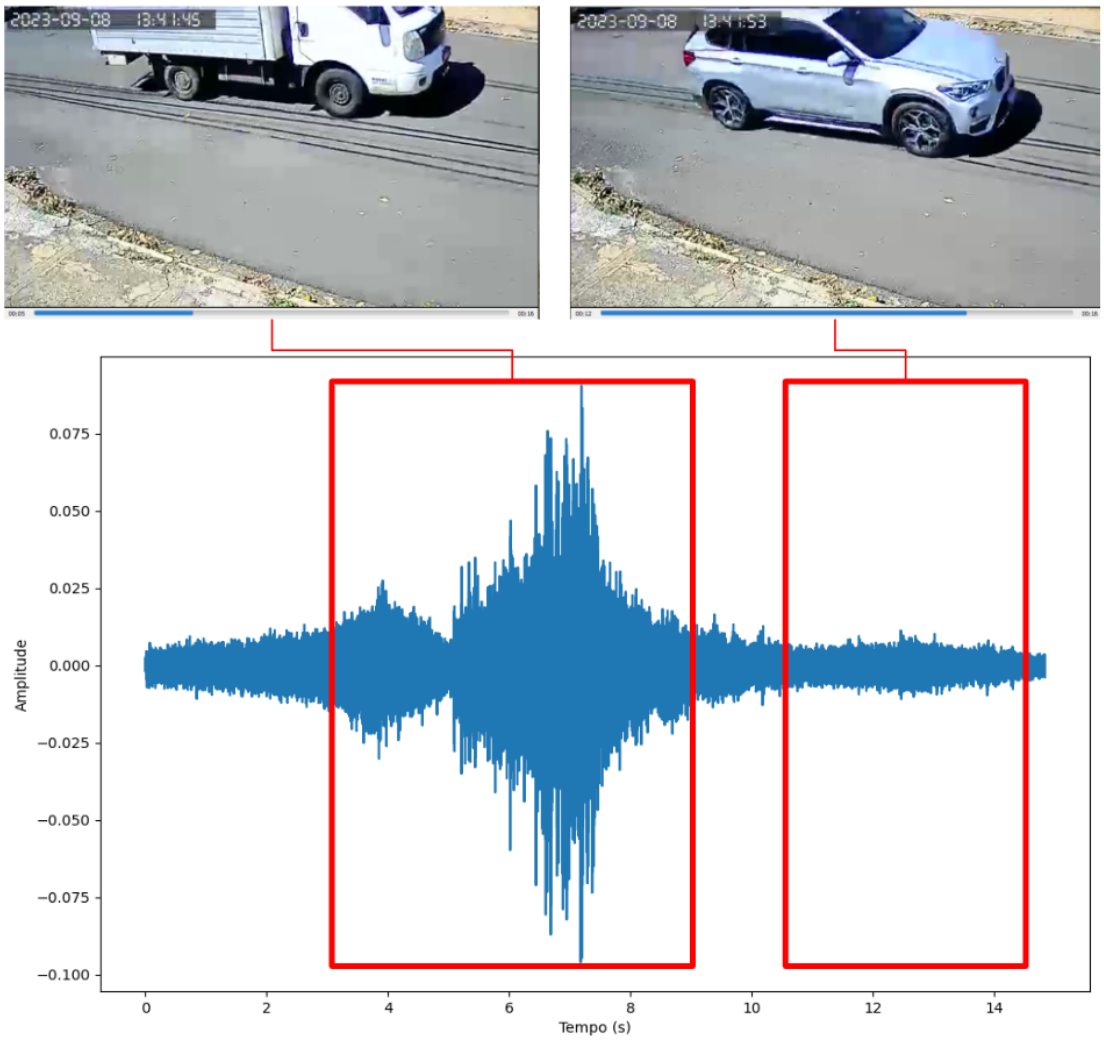

Figura 6: Amplitude do Microfone em função do tempo para a amostra de rua

com a qual o sinal cruza o eixo horizontal), desvio padrão do *zero-crossing rate* e treze MFCCs. As MFCCs, *Mel-frequency cepstral coefficients*, foram escolhidas, junto as outras features, para trazer uma análise no espectro da frequência do sinal, mostrando a composição dele. A escolha das MFCCs se deu observando sua característica logarítmica, e pelo uso difundido em projetos relacionados tanto à aprendizado de máquina [14] quanto veicular [15].

Iterando sobre todos os arquivos de áudio coletados, foi criado um arquivo CSV contendo em cada linha, uma detecção. Isto é, a primeira coluna com a *label* do tipo de veículo (ou "none"caso não haja veículos) e as demais com as *features* citadas.

6 Modelos de *Machine Learning*

Para a classificação das features geradas, foram escolhidos os modelos de *Random Forest* e *Convolutional Neural Network* (CNN), usando, para isso, as bibliotecas Sklearn [1] e Keras [2], respectivamente.

Dado a diferença muito grande na quantidade de detecções de carros e dos demais veículos,

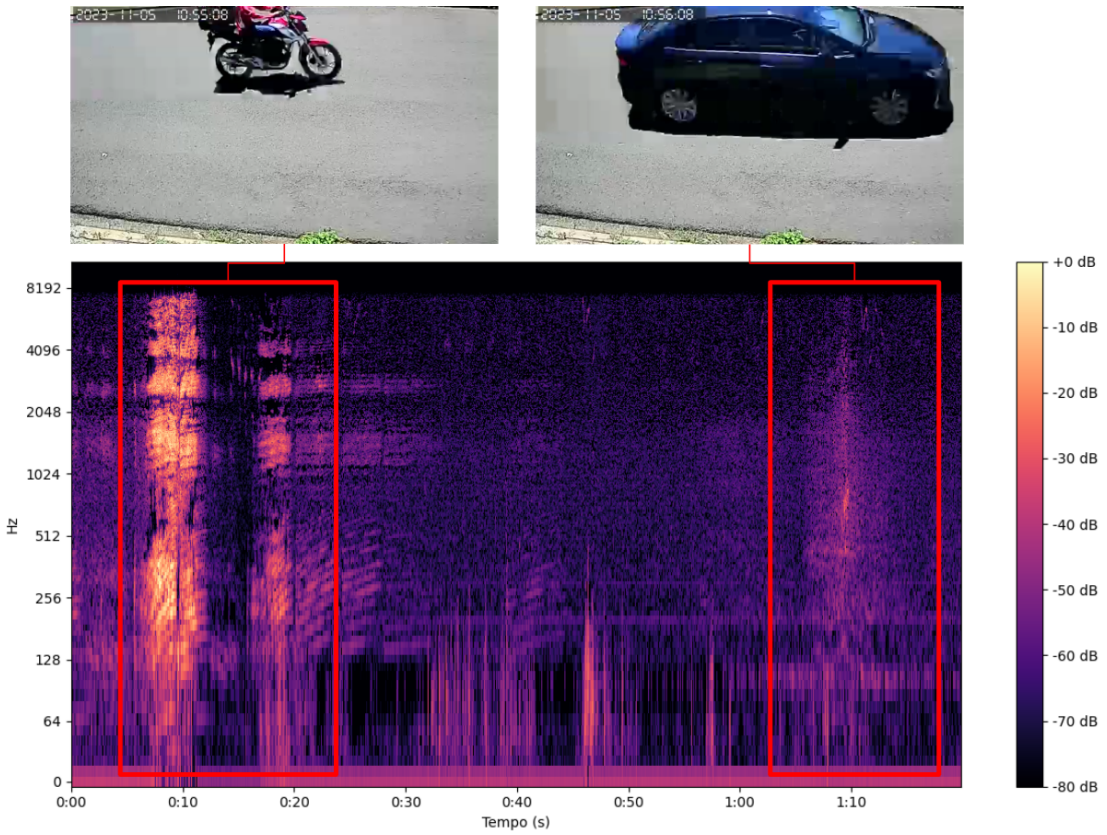

Figura 7: Espectrograma logarítmico da frequência para uma amostra da rua com a passagem de uma motocicleta e um carro

foram feitos dois treinamentos independentes de modelos, um agrupando todos os veículos frente à ausência de veículos e outro buscando a classificação entre carros, motocicletas e ausência de veículos, deixando os caminhões de lado graças a sua diminuta quantidade de dados. Além disso, para balancear as classes, o número de amostras da classe "none"foi configurado para ter aproximadamente a mesma quantidade da soma das demais.

6.1 Resultados obtidos na classificação de veículos

Para o modelo *Random Forest*, foi configurado como cem o número de estimadores, após testes. Nas tabelas 1 e 2 é possível observar a performance do modelo e na imagem 10 pode-se ver também as matrizes de confusão.

Tabela 1: Relatório Modelo *Ramdom Forest* para Classes Combinadas

	Precision	Recall	F1-Score	Support
None	0.97	0.94	0.96	720
Vehicle	0.94	0.97	0.96	698
Accuracy			0.96	1418
Macro Avg	0.96	0.96	0.96	1418
Weighted Avg	0.96	0.96	0.96	1418

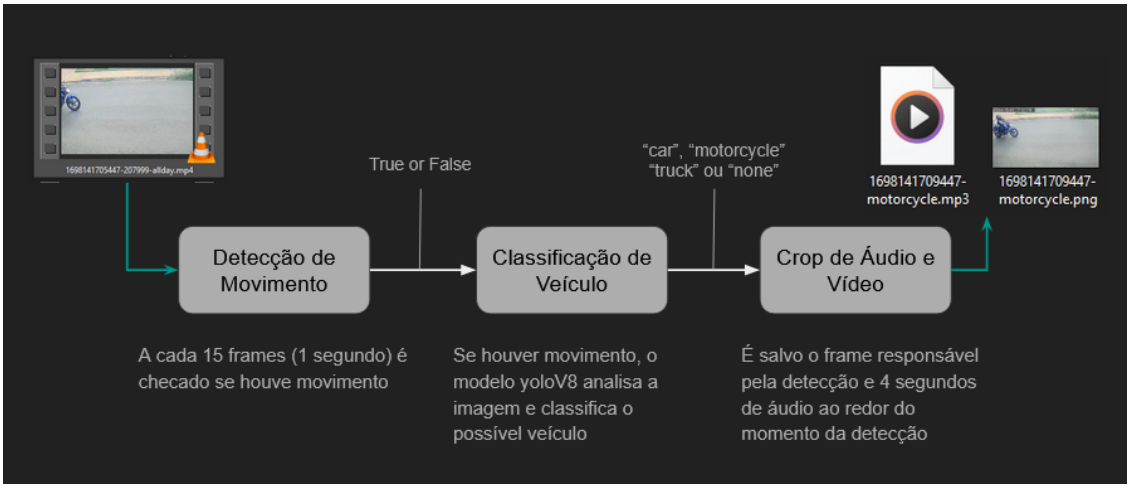

Figura 8: *Pipeline* de extração de áudios e imagens

Figura 9: Contraste entre as imagens capturadas durante o dia e durante a noite

Para a CNN, foram criadas 5 camadas internas sequenciais de dimensões 64, 128, 256, 64 e 32, usando como função de ativação a *Relu*. A camada final utilizou como função de ativação a *Softmax*. Dentre as diversas composições testadas, esta trouxe uma das melhores Acurácias sem um desbalanço muito grande nas demais métricas, além de um treinamento rápido. Nas tabelas 3 e 4 tem-se o relatório do modelo e na figura 11, temos as matrizes de confusão.

7 Desafios e Limitações

A execução deste projeto foi caracterizada por uma concentração significativa de esforços na etapa inicial de coleta e tratamento de dados, devido à decisão de não recorrer a bases de dados externas. Isso direcionou a atenção para a obtenção de informações diretamente do ambiente real, resultando em desafios notáveis. A análise do acelerômetro, discutida detalhadamente na seção 4.1, destacou a insuficiência da sensibilidade do sensor, representando uma das principais dificuldades iniciais enfrentadas.

Além disso, a coleta de dados da câmera apresentou desafios decorrentes das limitações impostas pelo fabricante, que restringiam a definição da imagem da câmera. A eficiência do processo de extração de áudio também foi um ponto crítico, levando cerca de duas horas para processar um dia de coleta, prolongando significativamente o ciclo de trabalho.

A etapa de limpeza dos dados coletados foi igualmente desafiadora, exigindo uma revisão mi-

Tabela 2: Relatório Modelo *Ramdom Forest* para Classes Separadas

	Precision	Recall	F1-Score	Support
Car	0.91	0.95	0.93	642
Motorcycle	0.84	0.53	0.65	86
None	0.95	0.95	0.95	591
Accuracy			0.92	1319
Macro Avg	0.90	0.81	0.84	1319
Weighted Avg	0.92	0.92	0.92	1319

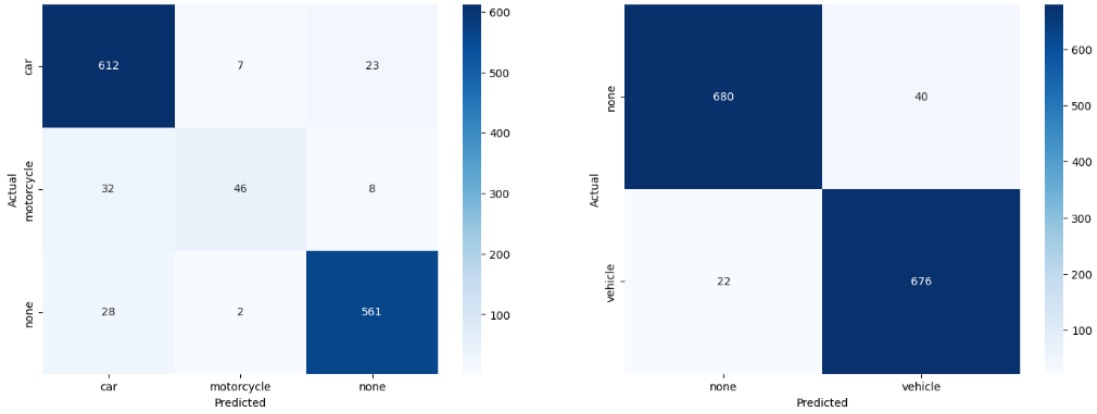

Figura 10: À esquerda, a matriz de confusão da *Random Forest* tendo como possíveis classes, "Carros", "Motos"e "Nenhum". À Direita, a matriz de confusão unindo todos os veículos frente a classe "Nenhum".

nuciosa para remover falhas na detecção, como veículos estacionados erroneamente classificados como em movimento, confusões entre caminhonetes e caminhões, e equívocos relacionados a bicicletas e motocicletas. Além disso, situações específicas, como a presença simultânea de motos e automóveis em um *frame*, exigiram a exclusão de dados para evitar a potencial confusão do modelo. Esses desafios, embora concentrados nas fases iniciais, destacam a complexidade intrínseca à obtenção e preparação de dados reais para treinamento de modelos de *machine learning*, reforçando a importância de uma abordagem meticulosa e adaptável ao longo do processo.

7.1 Limitações do estudo e possíveis melhorias

Existem claras limitações que influenciam diretamente a generalização e aplicabilidade dos resultados. Uma limitação evidente reside no cenário específico escolhido para a coleta de dados: uma rua residencial com um fluxo moderado de veículos. Este contexto, embora valioso para análises locais, impõe restrições à capacidade do modelo treinado em lidar com vias mais movimentadas, onde a diversidade de veículos e padrões de tráfego é significativamente maior. A adaptação do modelo a diferentes ambientes rodoviários pode requerer ajustes na metodologia de coleta, especialmente em vias com tráfego simultâneo de vários tipos de veículos.

Outra limitação diz respeito às condições meteorológicas variáveis, como chuvas intensas ou ventos fortes, que podem saturar a capacidade do microfone (como observado em alguns casos neste estudo), prejudicando sua eficácia tanto no momento de teste quanto de treinamento. Além disso, o estudo é sensível a condições ambientais diversas, como a presença de sons de obras ou

Tabela 3: Relatório de Treinamento de múltiplas classes para a CNN

	Precision	Recall	F1-Score	Support
Car	0.91	0.90	0.91	642
Motorcycle	0.59	0.51	0.55	86
None	0.92	0.95	0.93	591
Accuracy			0.90	1319
Macro Avg	0.81	0.79	0.80	1319
Weighted Avg	0.89	0.90	0.90	1319

Tabela 4: Relatório de Treinamento de classes combinadas para a CNN

	Precision	Recall	F1-Score	Support
None	0.94	0.95	0.95	720
Vehicle	0.95	0.94	0.94	698
Accuracy			0.94	1418
Macro Avg	0.94	0.94	0.94	1418
Weighted Avg	0.94	0.94	0.94	1418

música alta, que podem interferir nos dados coletados.

A diversidade de veículos também representa um desafio, especialmente em relação aos veículos elétricos, que apresentam perfis de áudio distintos. Motos, mesmo as convencionais, exibem uma variabilidade considerável na intensidade do ruído, adicionando complexidade à tarefa de detecção e classificação. Para superar essas limitações, melhorias potenciais incluem a expansão do conjunto de dados para abranger uma variedade mais ampla de cenários, condições climáticas e tipos de veículos.

Além disso, uma extração de *features* que leva em consideração o aspecto temporal da passagem de veículos, parece bastante promissora, sendo uma ideia futura de continuação desse trabalho. Ao analisar a Figura 7, é notável que, após a passagem da moto, surge um padrão distinto de eco a partir do segundo 20. Destaca-se também a observação de que, à medida que os veículos se aproximam e se afastam do microfone, manifesta-se, em muitos casos, um aumento seguido por uma diminuição na amplitude do áudio. Adicionalmente, em vias de maior velocidade, o efeito Doppler pode ser explorado, já sendo alvo de outros estudos[4]. Tais características oferecem potencial para aprimorar a precisão da detecção pelo microfone.

8 Conclusão

Em resumo, este trabalho abordou a detecção e classificação de veículos por meio de diversos sensores, destacando o uso de microfone e acelerômetro em conjunto com uma câmera. A ênfase na coleta e tratamento de dados reais, sem depender de bases externas, trouxe desafios iniciais significativos, especialmente na sensibilidade limitada do acelerômetro e nas restrições da coleta de dados da câmera. A implementação do modelo YOLOv8 revelou-se eficaz na detecção de veículos, com otimizações para lidar com o volume de dados, como paralelismo e detecção de movimento.

Contudo, limitações vinculadas ao cenário específico da rua residencial, às condições meteorológicas variáveis e à diversidade de veículos indicam que o modelo pode enfrentar desafios em ambientes mais complexos. Ainda assim, o estudo proporcionou informações valiosos sobre a viabilidade da abordagem multi sensores para a detecção veicular. Melhorias futuras incluem a expansão do conjunto de dados para cenários mais diversos, considerando diferentes condições climáticas e

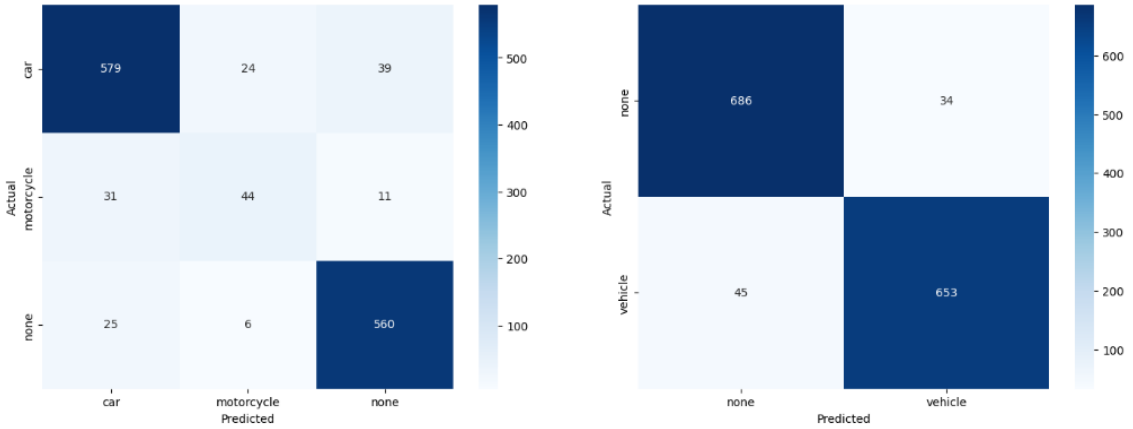

Figura 11: À esquerda, a matriz de confusão da CNN tendo como possíveis classes, "Carros", "Motos"e "Nenhum". À Direita, a matriz de confusão unindo todos os veículos frente a classes "Nenhum".

tipos de veículos, visando uma aplicação mais robusta em ambientes rodoviários diversos.

Destaca-se, entre os resultados, a eficácia média de 94% por parte da CNN no caso dos veículos combinados (detecção binária da passagem de um veículo).

Referências

[1] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[2] Francois Chollet et al. Keras, 2015.

[3] A. Feldbusch, H. Sadegh-Azar, and P. Agne. Vibration analysis using mobile devices (smartphones or tablets). *Procedia Engineering*, 199:2790–2795, 2017. X International Conference on Structural Dynamics, EURODYN 2017.

[4] J.F. Forren and D. Jaarsma. Traffic monitoring by tire noise. In *Proceedings of Conference on Intelligent Transportation Systems*, pages 177–182, 1997.

[5] Kratika Garg, Nirmala Ramakrishnan, Alok Prakash, and Thambipillai Srikanthan. Rapid and robust background modeling technique for low-cost road traffic surveillance systems. *IEEE Transactions on Intelligent Transportation Systems*, 21(5):2204–2215, 2020.

[6] Bo Jhen Huang, Jun-Wei Hsieh, and Chun Ming Tsai. Vehicle detection in hsuehshan tunnel using background subtraction and deep belief network. In Satoshi Tojo, Le Minh Nguyen, Ngoc Thanh Nguyen, and Bogdan Trawinski, editors, *Intelligent Information and Database Systems - 9th Asian Conference, ACIIDS 2017, Proceedings*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pages 217–226. Springer Verlag, 2017.

[7] N. K. Jain, R. K. Saini, and P. Mittal. A review on traffic monitoring system techniques. *Soft Computing: Theories and Applications*, pages 569–577, 2018.

[8] Anna Jakubczyk-Gałczyńska and Robert Jankowski. Traffic-induced vibrations. the impact on buildings and people. *The 9th International Conference "En-Vironmental engineering"*, 01 2014.

[9] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. YOLO by Ultralytics, January 2023.

[10] Jean-Philippe Jodoin, Guillaume-Alexandre Bilodeau, and Nicolas Saunier. Tracking all road users at multimodal urban traffic intersections. *IEEE Transactions on Intelligent Transportation Systems*, 17, 03 2016.

[11] Anurag Kanungo, Ayush Sharma, and Chetan Singla. Smart traffic lights switching and traffic density calculation using video processing. In *2014 Recent Advances in Engineering and Computational Sciences (RAECS)*, pages 1–6, 2014.

[12] B. McFee, Matt McVicar, Daniel Faronbi, Iran Roman, Matan Gover, Stefan Balke, Scott Seyfarth, Ayoub Malek, Colin Raffel, Vincent Lostanlen, Benjamin van Niekirk, Dana Lee, Frank Cwitkowitz, Frank Zalkow, Oriol Nieto, Dan Ellis, Jack Mason, Kyungyun Lee, Bea Steers, and Waldir Pimenta. librosa/librosa: 0.10.1, 2023.

[13] Henok Marie Shiferaw. Measuring traffic induced ground vibration using smartphone sensors for a first hand structural health monitoring. *Scientific African*, 11, 2021.

[14] Parwinder Singh. An approach to extract feature using mfcc. *IOSR Journal of Engineering*, 4:21–25, 08 2014.

[15] Lwin Thu and Htet Ne Oo. Acoustics-based vehicle classification system using mfcc and wavelet-based mfcc. 01 2018.

[16] RWTH Aachen University. Phyphox, 2023.

[17] Wikipedia contributors. Hsuehshan tunnel, n.d.

[18] Bing-Fei Wu, Chih-Chung Kao, Chih-Chun Liu, Chung-Jui Fan, and Chao-Jung Chen. The vision-based vehicle detection and incident detection system in hsueh-shan tunnel. *IEEE International Symposium on Industrial Electronics*, 06 2008.