

## Combinando YOLO com Visual Rhythm para Contagem de Veículos

Victor Nascimento Ribeiro

Nina S. T. Hirata

Instituto de Matemática e Estatística - Universidade de São Paulo

victor\_nascimento@usp.br

### Objetivos

A contagem de veículos é uma tarefa crucial para o controle do tráfego e gestão da infraestrutura de transporte. Este problema pode ser abordado por meio de métodos que se baseiam na detecção de veículos em vídeos. Consideramos cenários nos quais câmeras estáticas são utilizadas para capturar vídeos em vista superior de veículos movendo-se em uma única direção. Em vez de detectar e rastrear veículos em todos os frames, propomos a combinação do Visual Rhythm, uma técnica de geração de imagens espaço-temporais a partir de vídeos, com o YOLO, um modelo amplamente reconhecido para detecção de objetos. Isso nos permite processar seletivamente quadros relevantes, resultando em menor custo computacional.

### Métodos e Procedimentos

Definimos o problema de contagem de veículos como a tarefa de quantificar o número de veículos que atravessam uma linha designada (linha de contagem) no campo de visão da câmera. Isso pode ser realizado quadro a quadro detectando os veículos, rastreando-os e contando-os quando cruzam a linha. Para otimizar o processo, usamos o Visual Rhythm (VR) [1] para identificar quadros onde um objeto cruza a linha e aplicamos um modelo de detecção de veículos nesse quadro específico para o reconhecimento.

Seja  $f$  seja um vídeo com  $T$  quadros de tamanho  $M \times N$ . O VR de  $f$  é construído extraindo a linha de contagem de cada um dos

$T$  quadros e empilhando-os ao longo do eixo temporal. Isso resulta em uma imagem de tamanho  $T \times N$ , na qual cada linha da imagem VR representa a linha de contagem de um quadro, como ilustrado na Figura 1. Note que, por volta do quadro 20, um veículo atravessa a linha, deixando assim uma marca na VR.

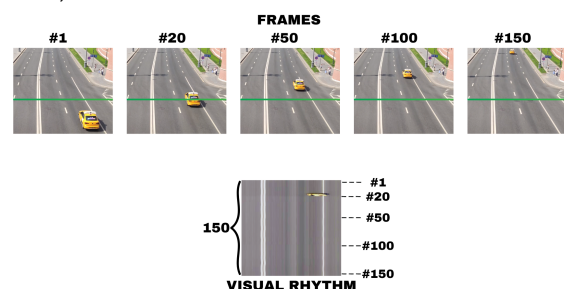


Figure 1: Geração do Visual Rhythm

Para a detecção tanto das marcas quanto dos veículos, utilizamos o YOLO (*You Only Look Once*) [2], um modelo de detecção de objetos baseado em aprendizado profundo.

O método proposto segue as etapas mostradas na Figura 2. Na primeira etapa (a), criamos uma imagem Visual Rhythm (VR) de  $T$  quadros consecutivos. Na etapa (b), usamos o YOLO para detectar marcas na imagem VR. Na etapa (c), extraímos o quadro correspondente a cada marca detectada. Posteriormente, é necessário confirmar se a marca em questão realmente corresponde a um veículo. Na etapa (d), o YOLO é usado novamente para detectar todos os veículos do quadro. Aquele que melhor se ajusta à marca em termos de tamanho e posição é selecionado na etapa (e) para atualizar a contagem de veículos por classe. Dividimos longas sequências de vídeo em

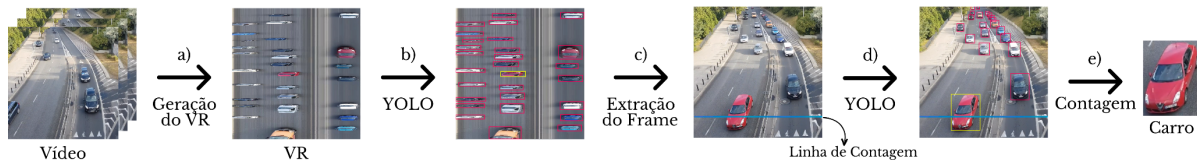


Figura 2: Passo a passo para a contagem de veículos baseada em VR.

segmentos não sobrepostos de comprimento  $T$  para gerenciar o tamanho da imagem de VR no processamento do YOLO.

mais baixas de 81,6% e 77,0%, respectivamente. Essas taxas se referem aos veículos detectados no conjunto de teste.

## Resultados

Para avaliar o desempenho do método proposto, selecionamos 4 vídeos disponíveis no YouTube. Veículos e marcas foram anotados e divididos em conjuntos de treinamento, validação e teste, garantindo que os mesmos carros não estivessem em mais de um subconjunto. Em seguida, refinamos o modelo pré-treinado *YOLOv8-small* para a detecção de veículos e marcas. Definimos a linha de contagem a uma altura de 120 de cada frame e utilizamos segmentos de vídeo com duração  $T = 900$  (30 seg) para gerar as imagens de VR.

Comparamos nossa abordagem com o sistema de contagem de veículos quadro a quadro da *Roboflow* [3], que utiliza *ByteTrack* para rastreamento de objetos em tempo real, em três vídeos. Mostramos os resultados na Tabela 1. Ambos os sistemas operaram em ambientes idênticos, utilizando os mesmos pesos de modelo. O método proposto, baseado em VR, mostrou ser cerca de três vezes mais rápido do que o baseado em rastreamento, uma vez que realiza a detecção de veículos somente em quadros selecionados.

Tabela 1: Acurácia de contagem (%) para cada vídeo

Sistema	Frame rate FPS	Vídeo 2	Vídeo 3	Vídeo 4
VR	186	100	98.9	98.5
Bytetrack	56	100	99.0	98.4

As taxas de acurácia de cada classe estão apresentadas na Tabela 2. Dentre as classes, 'carro' alcançou a maior acurácia, com 99,1%, enquanto 'Moto' e 'Caminhão' tiveram acurácia

Tabela 2: Acurácia de classificação (%) em cada vídeo usando o método VR

Carro	Ônibus	Moto	Pickup	Caminhão	Van
99.1	86.0	81.6	77.0	96.2	92.9

## Conclusões

O método proposto de contagem de veículos com base em VR é três vezes mais rápido do que o método baseado em rastreamento. Esse aumento de velocidade se deve à seleção eficiente apenas dos quadros de interesse, evitando a necessidade de realizar a detecção de veículos em todos os quadros do vídeo. Além disso, a precisão na contagem é comparável ao método baseado em rastreamento, o que demonstra sua eficácia. No entanto, há espaço para melhorias na identificação do tipo de veículo.

## Agradecimentos

MCTI (Brasil), Lei 8.248, PPI-Softex - TIC13 - 01245.010222/2022-44, FAPESP 2015/22308-2.

## Referências

- [1] S. Guimarães, M. Couprie, N. Leite, and D. A. Araújo, "A method for cut detection based on visual rhythm," in *Proceedings XIV Brazilian Symposium on Computer Graphics and Image Processing*, 2001, pp. 297–304.
- [2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CVPR*, pp. 779–788, 2016.
- [3] B. Dwyer, J. Nelson, J. Solawetz et al., "Roboflow (version 1.0) [software]," <https://roboflow.com>, 2022, computer vision.