Contagem e classificação de veículos com processamento de vídeos utilizando YOLOv3 e DeepSORT

Anderson Andrei Schwertner e Valner João Brusamarello

***Resumo*—O planejamento da instalação de sistemas e estruturas de controle e sinalização de rodovias depende de dados colhidos pelo Departamento Autônomo de Estradas de Rodagem (DAER). Atualmente esses dados são obtidos a partir do processamento de imagens capturadas por uma câmera temporariamente instalada em pontos solicitados. Essas imagens são pós-processadas para fornecer o número discriminado de cada classe de veículos bem como os sentidos de deslocamento, de forma refutar ou corroborar essas solicitações pela instalação ou modificação de alguma estrutura ou equipamento. Este trabalho apresenta o desenvolvimento de um sistema automático para realizar esse pós-processamento utilizando técnicas de processamento de imagens e deep learning.**

***Palavras-chave*—Rastreamento por detecção, Deep Learning, CNNs, YOLOv3 e DeepSORT.**

I. Introdução

Sistemas capazes de monitorar o tráfego de veículos têm um papel muito importante no desenvolvimento de cidades inteligentes. Dentre controle de tráfego, sinalização e infraestrutura, são muitas as contribuições que esses sistemas podem fornecer para tornar a malha de transporte mais eficiente e inteligente.

Dados sobre o fluxo de veículos em cada faixa, a direção das conversões realizadas e a classe dos veículos (carros, motocicletas, caminhões, etc) são informações que podem ajudar engenheiros de trânsito a melhor compreender a demanda e a distribuição do tráfego em múltiplas faixas, uma vez que esses dados são fundamentais no projeto de intersecções, sincronização de semáforos e também podem ser aplicados no desenvolvimento de outras estratégias para mitigação de congestionamento quando necessário.

Nesse âmbito, esta pesquisa busca suprir a demanda do Departamento Autônomo de Estradas de Rodagem do Rio Grande do Sul (DAER-RS) por um sistema automático para contagem, classificação e determinação da direção tomada pelos veículos em rodovias e intersecções. Essas informações são utilizadas pelos engenheiros do DAER-RS para refutar ou corroborar solicitações para instalação de semáforos, lombadas ou mesmo modificar estruturas já instaladas, e são atualmente obtidas após um processamento manual de imagens capturadas por uma câmera que é temporariamente instalada em cada local solicitado.

A. A. Schwertner, anderson.andrei@ufrgs.br, V. J. Brusamarello, valner.brusamarello@ufrgs.br, Tel +55-51-33084475

II. Revisão Bibliográfica

A obtenção automática dessas informações requer a utilização de técnicas de processamento de imagem capazes de detectar, classificar e rastrear os veículos ao longo de múltiplos frames. Especificamente, essas são tarefas comuns dentro da área de visão computacional e já foram testadas com técnicas clássicas de processamento de imagem como por exemplo filtros de partículas [1], operações morfológicas [2] e fluxo ótico [3], entretanto, a aplicação desses estudos foi sempre limitada pela até então inexistência de um método eficiente e confiável para a detecção dos veículos, consequentemente prejudicando o rastreamento no decorrer dos frames.

Foi a partir da popularização das Redes Neurais Convolucionais (CNNs), originalmente com a AlexNet [4] em 2012, que ocorreram grandes avanços na área de classificação e detecção de objetos em imagens.

A. Redes Neurais Convolucionais e a classificação de imagens

Redes Neurais Convolucionais fazem parte de um amplo grupo de técnicas conhecidas como Deep Learning e são arquitetadas a partir do empilhamento de múltiplas camadas de convolução, Max-Pooling e ativação (ReLU). A fórmula do seu êxito está nessa arquitetura, uma vez que permite capturar com sucesso as dependências locais e globais das imagens através de filtros que, em métodos primitivos, eram projetados de forma manual, e nas CNNs passaram a ser aprendidos no processo de treinamento da rede [5].

As metodologias baseadas em CNNs podem ser diretamente associadas a maior parte do progresso obtido na área nos últimos anos, o qual teve início quando [4] tornou-se a primeira CNN a vencer todas as outras propostas no ImageNet Challenge, obtendo uma taxa de erro de 16,4% na tarefa de classificação de imagens, comparada aos 26,2% do vencedor do ano anterior.

Desde então, todos os vencedores do ImageNet Challenge foram metodologias baseada em CNNs, apresentando anualmente um acelerado aumento na precisão e também na profundidade das CNNs. Enquanto que em 2012 a arquitetura vencedora foi construída com 8 camadas alcançando erro de 16,4%, a arquitetura vencedora em 2015 [6] utilizou 152 camadas atingindo erro de apenas 3,5%, sendo o estado da arte em 2020 de 1,3%.

Devido ao sucesso obtido na tarefa de classificação de imagens, essas arquiteturas de CNNs passaram também a ser utilizadas como blocos base (frequentemente referenciados como backbone) em algoritmos de detecção de objetos como [7] e [8], os quais, assim como na tarefa de classificação, observaram acelerado incremento na precisão com a utilização de CNNs como o classificador dos objetos detectados nas imagens.

B. Detecção de objetos

O processo de detecção de todos os veículos em uma imagem é diferente do simples caso de classificação, uma vez que o número de objetos e classes, neste caso veículos, em cada imagem é variável. Nessa perspectiva, existem dois grupos principais de algoritmos de detecção de objetos: detectores de estágio único e detectores de estágio duplo, sendo [7] e [8] respectivamente exemplos desses grupos.

A principal diferença entre as duas categorias é que os detectores de estágio duplo primeiramente encontram regiões de interesse, isto é, regiões em que possivelmente existe um objeto na imagem, para em seguida classificar as regiões separadamente, enquanto que detectores de estágio único estimam as bouding boxes (representação da posição e tamanho do objeto) e classificam os objetos simultaneamente.

Essa diferença faz com que detectores de estágio único sejam mais rápidos ao custo de precisão se comparados aos detectores de estágio duplo [6]. Entretanto, métodos mais modernos de estágio único como [7] conseguem obter precisão semelhante aos de estágio duplo com um tempo de inferência até 5x mais rápido, dependendo dos métodos, métricas e bases de dados comparados, fato que é evidenciado em [9], sendo [7] o mais rápido dentre os algoritmos analisados.

C. Rastreamento por Detecção

Uma das abordagens mais populares para o rastreamento de objetos em vídeos é a técnica conhecida como rastreamento por detecção. O rastreamento por detecção é um processo de duas etapas: um algoritmo de detecção de objetos é primeiramente utilizado para detectar os objetos presentes em cada frame; objetos que são então rastreados pela associação de objetos do frame atual com o frame anterior utilizando um algoritmo de rastreio [9].

O estágio de rastreio nesse método também pode ser visto como a solução de dois problemas distintos. Primeiramente as posições futuras dos objetos rastreados são estipuladas, usualmente utilizando Filtros de Kalman. Em seguida, objetos detectados em novos frames são associados com os objetos já rastreados em frames anteriores com base na previsão de posições futuras.

III. Metodologia

Dentre os múltiplos algoritmos de detecção e rastreamento de objetos disponíveis e desenvolvidos nos recentes anos, a escolha pela utilização do YOLOv3 [7] para detecção é embasada principalmente na boa relação de tempo de inferência versus taxa de erro reportada em [7] e [9], fato que permite obter precisão próxima ao estado da arte mas com taxa de quadros por segundo mais elevadas.

No que diz respeito ao DeepSORT [10], a escolha é fundamentada no seu bom desempenho para lidar com oclusão e reidentificação de objetos, problemas comuns nesta aplicação e que provocam erros no processo de contagem de veículos. Além disso, destaca-se que no trabalho realizado em [11], a utilização do DeepSORT foi a que obteve os melhores resultados ao rastrear carros e vans.

A. Contagem e direção

Através desses dois algoritmos são obtidas as informações utilizadas para a contagem e determinação da direção dos veículos. Durante o processo de detecção e rastreamento, são armazenadas as coordenadas de cada veículo no primeiro e último frame em que são detectados no vídeo. Esse conjunto de coordenadas é então analisado como um problema de clusterização que quando resolvido gera regiões da imagem onde carros surgem e desaparecem, sendo estas as direções possíveis para cada veículo detectado.

IV. Cronograma

	2020				2021
	1º Tri	2º Tri	3º Tri	4º Tri	1º Tri
Revisão bibliográfica	✓	✓	X		
Filtragem e preparação do Banco de Dados	✓	✓			
Estudo de Python e Deep Learning	✓	✓			
Detector de veículos		✓			
Rastreador de veículos, Contagem e Direção			X	X	
Obtenção e Análise dos resultados			X	X	
Escrita de artigo para publicação			X		
Escrita da dissertação		✓	X	X	X
Defesa					X

Referências

[1] C. Bouvié, J. Scharcanski, P. Barcellos, and F. L. Escouto, "Tracking and counting vehicles in traffic video sequences using particle filtering." IEEE, 2013, pp. 812–815.

[2] H. T. P. Ranga, M. R. kiran, S. R. shekar, and S. K. N. kumar, "Vehicle detection and classification based on morphological technique," in *International Conference on Signal and Image Processing*, 2010.

[3] A. Abdagic, O. Tanovic, A. Aksamovic, and S. Huseinbegovic, "Counting traffic using optical flow algorithm on video footage of a complex crossroad," in *Proceedings ELMAR-2010*, Sep. 2010, pp. 41–45.

[4] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Information Processing Systems*, vol. 25, 01 2012.

[5] *Machine Learning: Methods and Applications to Brain Disorders*. ACADEMIC PR INC, 2019. [Online]. Available: https://www.ebook.de/de/product/36978686/machine_learning_methods_and_applications_to_brain_disorders.html

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition."

[7] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks."

[9] A. Nyström, "Evaluation of Multiple Object Tracking in Surveillance Video ," Master's thesis, Linköping University, Sweden, 2019.

[10] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric."

[11] *Computer Vision - ECCV 2018 Workshops*. Springer-Verlag GmbH, 2019. [Online]. Available: https://www.ebook.de/de/product/34982719/computer_vision_eccv_2018_workshops.html