

APLICAÇÃO DE APRENDIZAGEM PROFUNDA PARA DETECÇÃO AUTOMÁTICA DE SINALIZAÇÃO VERTICAL EM RODOVIAS

Luciana da Silva Nascimento

Aline Calheiros Espíndola

Ricardo Venescau de Oliveira Almeida

Ernesto Ferreira Nobre Júnior

Universidade Federal do Ceará

Programa de Pós-graduação em Engenharia de Transportes

RESUMO

A sinalização é de extrema importância para a redução dos números de acidentes que ocorrem nas rodovias. Para manter a atenção dos usuários, no entanto, é necessária uma boa visibilidade dessa sinalização. Porém, para ter informações sobre a qualidade da sinalização vertical, é necessário o inventário desses elementos, processo que pode se mostrar bastante moroso e oneroso. Sendo assim, algumas técnicas de visão computacional, como o algoritmo You Only Look Once (YOLO), podem auxiliar neste processo de coleta de informações. O objetivo da presente pesquisa é apresentar uma proposta de detecção automática dos elementos de sinalização vertical, analisando cenários com diferentes tamanhos de imagem e diferentes versões do YOLO. A Convolutional Neural Network (CNN) que apresentou os melhores resultados foi aquela que considerou o Yolov3, com tamanho de imagem de 416x416 pixels e 5000 iterações, cujas métricas obtidas foram: F1-score de 0,87, IoU médio de 63,22% e mAP@0,50 de 86,0%.

ABSTRACT

Signage is extremely important to reduce the number of accidents that occur on the highways. To keep users' attention, however, good visibility of this sign is needed. However, to obtain information about the quality of vertical signage, it is necessary to inventory elements, a process that can prove to be quite time-consuming and costly. Therefore, some computer vision techniques, such as the You Only Look Once (YOLO) algorithm, can help in this information gathering process. The objective of this research is to present a proposal for automatic detection of vertical signaling elements, analyzing scenarios with different image sizes and different versions of YOLO. The Convolutional Neural Network (CNN) that presents the best results was the one considered Yolov3, with an image size of 416x416 pixels and 5000 iterations, which causes metrics such as: F1-score of 0.87, an average IoU of 63.22% and mAP@0,50 of 86.0%.

1. INTRODUÇÃO

Segundo o relatório da Organização Mundial de Saúde (2015), 1,25 milhão de pessoas morrem anualmente no mundo como vítimas de acidentes de trânsito, sendo 42 mil vítimas no Brasil. No estado do Ceará, segundo dados disponibilizados pela polícia rodoviária, nas rodovias federais foram registrados 1.696 acidentes em 2018, sendo 1120 com vítimas feridas, e 164 fatais. Essa situação ainda é mais acentuada nas rodovias estaduais, com 2.491 acidentes. Segundo o Retrato da Segurança Viária no Brasil (2017), os impactos dos acidentes de trânsito são percebidos na rede de saúde pública, contribuindo para a sobrecarga de leitos hospitalares, perda de produtividade econômica e ônus aos cofres públicos. Segundo a CNT (2019), em 2018, o prejuízo total para a economia brasileira em decorrência dos acidentes foi de aproximadamente R\$ 9,73 bilhões.

Embora o fator humano seja considerado a principal causa dos acidentes, nem sempre as medidas de mitigação associadas a este componente apresentam melhor benefício/custo. Segundo Nodari (2003) e Tamayo (2010), o bom desempenho do motorista depende da combinação de características como geometria da via, sinalização horizontal e vertical e da clara definição de prioridades de circulação.

Segundo o DNER (1998), é possível atuar no combate à redução de acidentes de duas formas:

implantando soluções de grande porte ou utilizando soluções de baixo custo. As soluções de grande porte têm como objetivo eliminar completamente algum fator gerador de acidente através de mudanças radicais nas características das vias; e as soluções de baixo custo objetivam reduzir os riscos, o que não necessariamente significa que eles serão eliminados.

Segundo o DNER (1988), a experiência mundial mostra que soluções de baixo custo vêm apresentando respostas significativas, principalmente em países em desenvolvimento. Visto isto, uma das soluções de baixo custo é uma melhora na sinalização, mais especificamente na sinalização vertical, enfoque da presente pesquisa. A sinalização vertical tem por objetivo advertir os usuários de perigos, isto é, preveni-los da existência de curvas, pista derrapante, declives, dentre outras ocorrências.

Observando os dados da qualidade da sinalização vertical no Brasil fornecidos pela CNT (2019), 52.417 km apresentam problema de sinalização, com 22% em estado péssimo e ruim, ampliando os riscos de acidentes nas rodovias. Sendo assim, é notória a importância de conhecer a qualidade dos elementos rodoviários, para que os gestores possam elaborar um plano de tomadas de decisões que permita atingir uma segurança viária adequada. Para isso, é necessário inventariar e georreferenciar os elementos rodoviários na malha administrada. No entanto, segundo Santos (2016), esta inventariação é um processo moroso, devido à extensa malha rodoviária, que necessita ser percorrida para localizar, identificar e georreferenciar os elementos de interesse.

Frente a estas dificuldades e à importância do assunto em questão para a sociedade como um todo, atualmente vem crescendo o número de pesquisadores interessados na automação dos levantamentos dos elementos rodoviários. A automação pode ser realizada com diversas técnicas de visão computacional e inteligência artificial, sendo uma tendência atual a utilização de técnicas de detecção de objetos com o auxílio do algoritmo You Only Look Once (YOLO).

Sendo assim, o objetivo da presente pesquisa é apresentar uma proposta de detecção automática dos elementos de sinalização vertical, identificando três classes: placas regulamentares, advertência e indicativas, analisando cenários com diferentes tamanhos de imagem e versões do YOLO.

2. REVISÃO DE LITERATURA

2.1. Sinalização vertical

A sinalização é padronizada em todo o território, devendo atender aos princípios norteadores definidos pelo Manual Brasileiro de Sinalização de Trânsito. Tais princípios são: legalidade, suficiência, padronização, clareza, precisão, visibilidade, legibilidade, manutenção e conservação. Para atendimento aos mesmos, é necessário que haja uma implantação adequada e manutenção permanente. A sinalização pode ser subdividida em horizontal, vertical e dispositivos auxiliares, sendo o enfoque do presente trabalho a sinalização vertical. Segundo o Manual de Sinalização do DNIT (2010), tal sinalização tem por finalidade a regulamentação do uso da via, advertência para situações potencialmente perigosas ou problemáticas e o fornecimento de indicações, orientações e mensagens educativas.

Segundo o Código de Trânsito Brasileiro (2010), a sinalização vertical é classificada de acordo com sua função, compreendendo a sinalização de regulamentação, de advertência e de indicação. A sinalização de regulamentação tem por objetivo informar aos usuários as

condições, proibições, obrigações ou restrições no uso das vias; forma padrão é circular, apresentando as cores vermelha, preta e branca. A sinalização de advertência tem por finalidade alertar os usuários para condições potencialmente perigosas; forma padrão é quadrada, com uma das diagonais na posição vertical, apresentando as cores amarela e preta. E a sinalização de indicação tem por finalidade identificar vias e locais de interesse, podendo ter como função a educação do usuário; pode apresentar diferentes formatos, e suas cores são azul, marrom, branca, verde e preta. A Figura 1 exemplifica algumas placas dessas categorias.

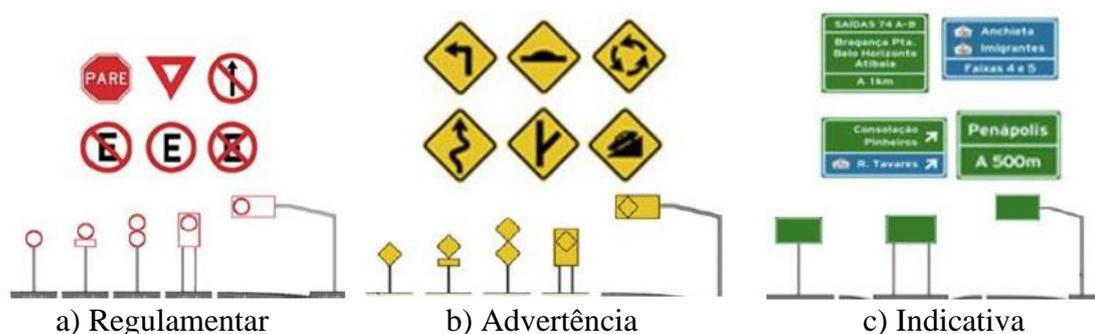


Figura 1: Exemplos de sinalização vertical regulamentar, de advertência e indicativa. Fonte: Código de trânsito Brasileiro (1997)

2.2. Visão computacional e Aprendizagem Profunda

2.2.1. Aprendizagem Profunda

Segundo Voinigescu (2017), aprendizagem profunda (*deep learning*) é uma subcategoria de *machine learning*, a qual emprega técnicas adequadas para a construção de redes neurais de grandes dimensões, capazes de processar elevado volume de dados de forma eficiente. O termo “profunda” refere-se ao fato de estas serem construídas com o emprego de um maior número de camadas em relação às redes neurais tradicionais. Estas camadas adicionais são responsáveis por extrair/representar a informação a partir de diferentes graus de abstração. A Figura 2 esquematiza a rede de neurônios com suas camadas intermediárias e os pesos que são aprimorados a cada iteração.

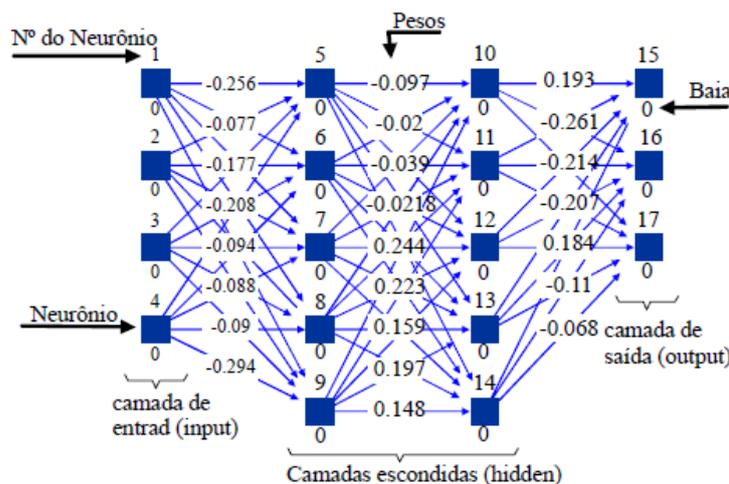


Figura 2: Rede de neurônios. Fonte: Gleriani (2004)

Segundo Ikram (2019), recursos de aprendizagem profunda em vários níveis de abstração permitem que um sistema aprenda funções complexas mapeando a entrada para a saída

diretamente dos dados, sem depender de recursos criados por humanos. As tecnologias de aprendizagem profunda obtiveram sucesso em várias tarefas de visão computacional, como classificação e segmentação de imagens e detecção de objetos.

2.2.2. Redes Neurais Convolucionais

A Rede Neural Convolutiva (*Convolutional Neural Network - CNN*), é uma abordagem de aprendizado profundo que pode captar uma imagem de entrada, atribuir importância (pesos e vieses) a vários aspectos/objetos da imagem e capaz de diferenciar um do outro. Além disso, é capaz de capturar com sucesso as dependências espaciais e temporais em uma imagem através da aplicação de filtros. Isto possibilita um melhor ajuste ao conjunto de dados da imagem devido à redução no número de parâmetros envolvidos e à capacidade de reutilização dos pesos. Na Figura 3 é possível observar a arquitetura de uma CNN.

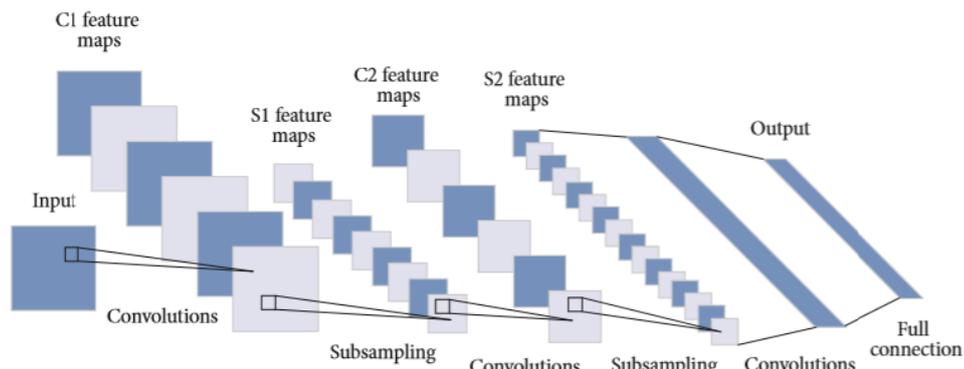


Figura 3: Arquitetura de uma CNN. Fonte: Xie, Zhang e Bai (2017)

2.2.3. You Only Look Once (YOLO)

YOLO é um método de detecção de objetos voltado para processamento em tempo real, que se baseia na arquitetura das redes neurais convolucionais, possuindo grande capacidade de generalização (REDMON *et al.* 2016). O procedimento realizado pelo YOLO para detecção de objetos segue os seguintes passos, de acordo com Hollerweger (2019):

- i. A imagem é dividida em grid, como demonstrado na Figura 4a.
- ii. Cada célula é responsável pela predição de caixas de seleção. Tais caixas podem ter qualquer tamanho, e correspondem à área onde a rede prevê a existência do objeto. Neste passo, são obtidos *scores* de confiança (probabilidade de existência de objeto) para cada uma das caixas. Essas caixas são ilustradas na Figura 4b, em que caixas de espessura mais grossas representam maior probabilidade de existência de objeto.
- iii. Para cada uma das caixas de seleção, a rede faz a predição de uma classe, o *score* de confiança da caixa contenha um objeto é combinado com um novo *score* que representa a probabilidade de acerto do objeto classificado. O resultado deste passo pode ser visualizado na Figura 4c, em que caixas de diferentes *scores* representam diferentes classes de objetos.
- iv. Por fim, partes das caixas possuem um *score* baixo e são eliminadas, caso o *score* seja menor que o limite estabelecido, comumente de 30%. Após tal operação, são obtidas as caixas com maior probabilidade de conter objetos classificados corretamente pela rede. O resultado pode ser visualizado na Figura 4d.

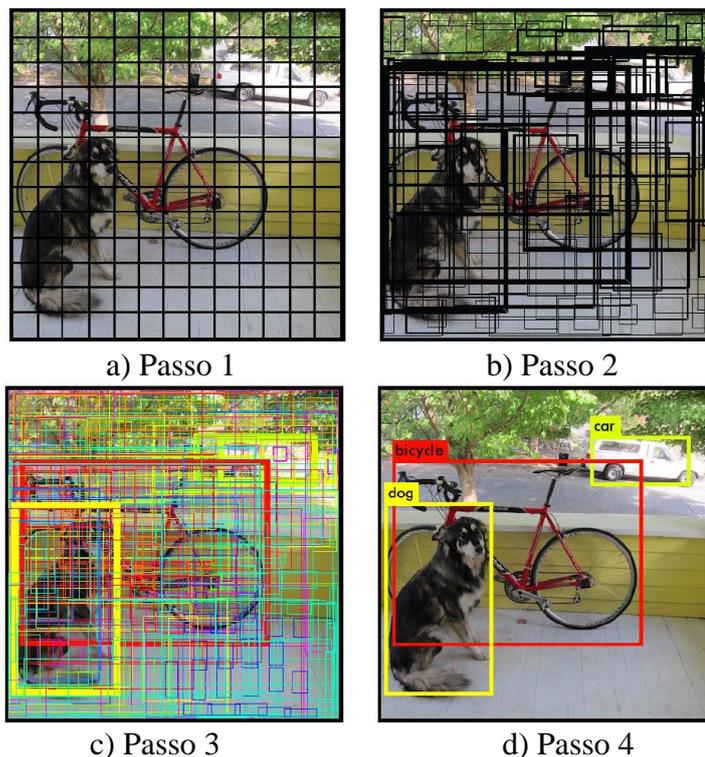


Figura 4: Estratégias YOLO. Fonte: Hollerweger (2019)

Segundo Zhang *et al.* (2018), o YOLO apresenta três grandes vantagens: em primeiro lugar, rompe o limite de velocidade máxima da CNN e atinge um excelente equilíbrio entre velocidade e precisão; em segundo lugar, raciocina globalmente e codifica informações contextuais sobre a imagem, sendo menos provável que preveja falsos positivos em segundo plano; e em terceiro lugar, aprende representações gerais de objetos, ultrapassando outros métodos de detecção, incluindo DPM e R-CNN por uma ampla margem, ao generalizar de imagens naturais para outros domínios como a arte.

Ainda segundo Zhang *et al.* (2018), o YOLO apresenta algumas limitações, como imposição de fortes restrições espaciais nas previsões da caixa delimitadora como reconhecer pequenos objetos em grupo; dificuldade para generalizar objetos novos ou incomuns em relações de aspecto ou configurações; trata a função de perda do YOLO de forma semelhante tanto em pequenas caixas delimitadoras como em caixas grandes.

Segundo Mendes e González (2020), YOLO tem evoluído em suas diferentes versões, incluindo YOLOv3 e YOLOv4, as últimas versões estáveis. A diferença entre as versões é que YOLOv3 usa redes de pirâmide características (FPN) para detecção de objetos, enquanto YOLOv4 usa PANet como um método de agregação de parâmetros para diferentes níveis de detecção junto com um aumento na precisão média (AP) e quadros por segundo (FPS), um recurso que torna no geral a precisão do YOLOv4 maior do que a do YOLOv3.

Apesar de existir YOLOv5, optou-se por não utilizar na presente pesquisa, tendo em vista que os desenvolvedores do YOLOv5 não são os mesmos das versões anteriores, não possuindo nenhuma ligação com as versões iniciais do YOLO. Como também, os autores não apresentaram artigo científico explicando os detalhes da estrutura e método utilizados na

detecção. As informações da versão 5 são escassas, bem como os testes realizados para verificar a real melhoria no desempenho.

3. MÉTODO

3.1. Conjunto de Dados

As imagens dos conjuntos de dados foram obtidas de várias rodovias estaduais localizadas no estado do Ceará, Brasil. Estas imagens foram capturadas pela câmera Garmin VIRB Ultra30 Action, acoplada em um veículo de pesquisa, produzindo dimensões de 1920x1080.

O enfoque da pesquisa são as sinalizações verticais, logo as classes utilizadas foram placas regulamentares, de advertência e indicativas. Foram selecionadas 1000 imagens e a rotulação foi realizada com o auxílio do software labelImg. Em algumas imagens foram rotuladas mais de uma placa, conforme Figura 5, fato este que produziu um total de 1099 rótulos, sendo 386 de placas regulamentares, 374 de placas de advertência e 339 de placas de indicação. Durante o treinamento da Yolo, 80% das imagens foram utilizadas para o treinamento e 20% para validação, escolhidas aleatoriamente.

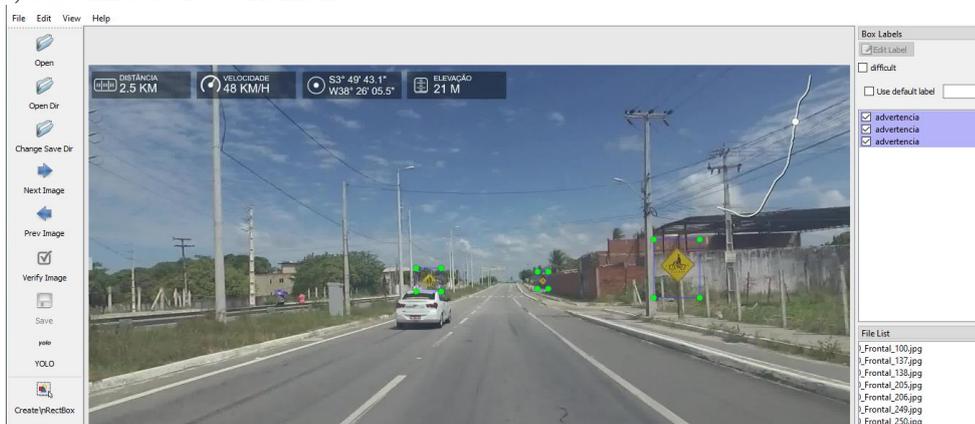


Figura 5: Rotulação de sinalização vertical nas rodovias

3.2. Procedimento Proposto

O treinamento da CNN tem um grande consumo de energia computacional. Portanto, é necessário usar GPUs que possam processar dados em paralelo. Para isso, os recursos de computação do Google Colaboratory (Colab) foram usados para conduzir o treinamento da CNN. Devido à natureza do Google Colab, cada treinamento é realizado em uma GPU alocada aleatoriamente. Todos os quatro modelos foram treinados com a GPU Tesla T4 – 13,6 GB.

O treinamento é um procedimento que envolve otimização matemática, envolvendo uma função de custo, chamada função de perda, para que os pesos e vieses da rede neural possam ser ajustados automaticamente. Neste processo, os dados de treinamento com rótulos esperados são apresentados ao procedimento de otimização para encontrar uma função de perda mínima global.

Os modelos são baseados na aprendizagem de transferência e os backbones usados para o treinamento foram CSPDarknet53 para Yolov4 e Darknet-53 para Yolov3. Os tamanhos das imagens testados foram 224x224 e 416x416 pixels, o tamanho do lote de 64 e subdivisões de 16. A camada final de classificação apresenta cruzamento binário-entropia como função de ativação (sigmóide) para as previsões de classe. Para todos os modelos desenvolvidos foram utilizados os mesmos parâmetros de otimização: *momentum* utilizado foi de 0,949, *decay* de

0,0005 e *learning rate* (LR) 0,001.

Como é mostrado em Wang e Liao (2020), o número de interações recomendados pelo desenvolvedor Yolo é de 2.000 para cada classe, não podendo ser inferior a 6.000 iterações. Durante o desenvolvimento desta pesquisa, foi analisada a expansão progressiva do número de iterações de treinamento, 1000 de intervalo, com um valor mínimo de 1000 a um valor máximo de 6000. Foram utilizados cinco métricas diferentes, a saber: *intersect over union* (IOU, em português é interseção sobre a união), *precision* (precisão), *recall* (sensibilidade), *F1-score*, *average precision* (AP, em português é precisão média) e *mean average precision* (mAP, em português é medias das precisões de todas as métricas).

A IOU pode ser calculada como a área de intersecção dividida sobre a área de união de duas caixas, portanto, a IOU deve estar entre 0 e 1. Ao prever caixas delimitadoras, é preciso encontrar o IOU entre a caixa delimitadora prevista e a caixa de verdade fundamental para ser ~ 1. A equação 1 apresenta o cálculo do IoU.

$$IoU = \frac{\text{Área de interseção}}{\text{Área da União}} \quad (1)$$

O parâmetro *precision* pode definir a precisão como a proporção de verdadeiros positivos - (previsões verdadeiras) e o número total de positivos previstos (previsões totais – verdadeiros positivos VP mais falsos positivos FP). A equação 2 apresenta o cálculo de *precision*.

$$Precision = \frac{VP}{VP + FP} \quad (2)$$

Recall é a proporção de verdadeiros positivos (previsões verdadeiras) e o total de verdadeiros positivos (verdadeiros positivos VP mais falsos negativos FN)). A fórmula equação 2 é apresenta o cálculo de *recall*.

$$Recall = \frac{VP}{VP + FN} \quad (3)$$

O parâmetro F1-score é uma média harmônica dos parâmetros *precision* e *recall*, sendo representada pela equação 4. A média harmônica dá mais peso aos valores baixos, sendo assim, só será obtida uma pontuação alta de F1-score se tanto o *recall* quando as precisões forem altas.

$$F_1 = \frac{VP}{VP + \frac{FN+FP}{2}} \quad (4)$$

Average precision (AP) e *mean average precision* (mAP): Uma breve definição para a precisão média é a área sob a curva de recuperação de precisão. AP combina precisão e recall. Tem um valor entre 0 e 1 (quanto maior, melhor). Para obter AP = 1, é necessário que a precisão e a recuperação sejam iguais a 1. O mAP é a média do AP calculado para todas as classes.

4. RESULTADOS

Os resultados obtidos para os diferentes cenários encontram-se dispostos na Tabela 1, sendo considerado performance satisfatória com valores superiores a 85% para as métricas de classificação (*Precision*, *Recall*, *F1-score*, mAP@ 0.50) e em torno de 65% na métrica de localização (IoU). Comparando inicialmente a dimensão das imagens, pode-se observar que o aumento do tamanho das imagens resultou, em geral, em melhores resultados métricos, tanto no Yolov3 quanto no Yolov4, resultado este que era esperado.

Tabela 1: Resultado das métricas dos quatro treinamentos realizados

Iterações	AP			Precision	Recall	F1-Score	Average IoU	mAP@0.50
	Regulamentar	Advertência	Indicativa					
Modelo YOLOv3 - Dimensão 224x224								
1000	70,91%	44,38%	64,54%	0,73	0,48	0,58	48,79%	59,94%
2000	68,90%	66,47%	98,43%	0,82	0,75	0,78	57,65%	77,93%
3000	73,37%	60,06%	94,70%	0,79	0,80	0,79	56,00%	76,04%
4000	72,07%	70,81%	94,52%	0,84	0,80	0,82	61,10%	79,13%
5000	71,81%	76,86%	90,31%	0,82	0,82	0,82	59,78%	79,66%
6000	72,66%	75,69%	94,70%	0,83	0,82	0,82	60,75%	81,02%
Modelo YOLOv3 - Dimensão 416x416								
1000	68,25%	56,85%	70,79%	0,64	0,70	0,67	41,08%	65,30%
2000	82,55%	70,62%	86,31%	0,86	0,83	0,85	60,12%	79,82%
3000	69,72%	85,69%	91,30%	0,84	0,82	0,83	62,58%	82,24%
4000	80,19%	65,18%	94,79%	0,91	0,80	0,85	64,45%	80,05%
5000	71,64%	93,21%	93,15%	0,88	0,87	0,87	63,22%	86,00%
6000	70,22%	87,58%	94,52%	0,88	0,85	0,86	64,07%	84,11%
Modelo YOLOv4 - Dimensão 224x224								
1000	68,58%	66,44%	79,44%	0,6	0,82	0,69	42,06%	71,49%
2000	52,80%	59,33%	89,45%	0,65	0,78	0,71	45,95%	67,19%
3000	75,63%	70,12%	91,30%	0,81	0,83	0,82	60,44%	79,02%
4000	77,81%	79,25%	92,63%	0,87	0,88	0,88	64,02%	83,23%
5000	*	*	*	*	*	*	*	*
6000	73,53%	79,31%	93,19%	0,85	0,85	0,85	63,72%	82,01%
Modelo YOLOv4 - Dimensão 416x416								
1000	66,50%	82,51%	78,30%	0,63	0,87	0,73	42,60%	75,77%
2000	73,62%	90,74%	86,09%	0,83	0,83	0,83	59,90%	83,48%
3000	74,08%	79,57%	88,08%	0,81	0,87	0,84	60,33%	80,58%
4000	60,68%	99,42%	79,70%	0,85	0,87	0,86	61,29%	79,93%
5000	73,40%	74,44%	85,87%	0,82	0,85	0,84	60,84%	77,90%
6000	71,27%	72,44%	91,67%	0,85	0,87	0,86	63,38%	78,46%

* Durante o treinamento do YOLOv4, dimensão 224x224, não foi possível exibir os resultados referentes à iteração 5000 e por isso as informações dessa linha encontram-se vazias.

O CNN que apresentou os melhores resultados baseou-se no YOLOv3, com tamanho de imagem de 416x416 pixels e 5000 iterações, e obteve F1-score de 0,87, IoU médio de 63,22% e mAP@0,50 de 86,0%. O segundo melhor modelo baseou-se no YOLOv3, 416x416 pixels, com 6000 iterações. E o terceiro foi a base do YOLOv4, dimensão 224x224 pixels, com 4000 iterações, resultado inesperado visto que as imagens com maiores dimensões apresentaram resultados melhores métricas para as demais iterações.

Em geral, a expansão das iterações levou a uma melhoria no IoU e F1-Score, mas não necessariamente nas outras métricas. A análise dos resultados do mAP não mostrou uma tendência específica quanto ao número de interações e ganho de desempenho.

Observando especificamente a métrica AP, pode-se perceber que as placas indicativas são identificadas com mais facilidade se comparado as demais classes, enquanto as placas regulamentares são levemente mais difíceis. Tal fato pode ser explicado porque o YOLO tem dificuldade para identificar objetos pequenos e, no dataset, há tanto placas distantes da câmera quanto placas mais próximas, e, por consequência, maiores. Como na maioria das imagens do banco de dados de treinamento as placas são maiores, é justificável o erro nas placas mais distantes.

As Figuras 6, 7, 8 e 9 mostram exemplos de detecção realizada pelo melhor modelo: Modelo YOLOv3 - 416x416 pixels treinados com 5000 iterações. Na Figura 6, o modelo conseguiu

detectar uma placa regulamentar com uma precisão de 98%; na Figura 7, conseguiu detectar uma placa de advertência com uma precisão de 100% e na Figura 8, conseguiu detectar uma placa indicativa com uma precisão de 99%. Na Figura 9, observa-se que o modelo conseguiu identificar mais de uma placa em uma única imagem e, além disso, verificou-se que ele consegue detectar placas que estão levemente distantes da câmera, a exemplo da placa de advertência ao fundo detectada com precisão de 71%.



Figura 6: Placa regulamentar identificada automaticamente utilizando-se o Modelo Yolov3-416x416 – 5000 iterações.



Figura 7: Placa de advertência identificada automaticamente utilizando-se o Modelo Yolov3-416x416 – 5000 iterações.



Figura 8: Placa indicativa identificada automaticamente utilizando-se o Modelo Yolov3-416x416 – 5000 iterações.



Figura 9: Sinalização vertical identificada automaticamente utilizando-se o Modelo Yolov3-416x416 – 5000 iterações.

Apresentar um modelo que permita a identificação automatizada da sinalização vertical, por mais que não apresente métricas superiores a 90%, é de grande vantagem, já que possibilita um ganho de tempo e de mão de obra durante processo de elaboração do inventário da rodovia.

5. CONCLUSÃO

Com a pesquisa, foi possível verificar que a arquitetura Yolo conseguiu detectar as placas de sinalização vertical de forma satisfatória, com o modelo Yolov3 - 416x416 mostrando os melhores resultados de métricas combinadas. Apesar das imagens com 416x416 pixels apresentarem melhores resultados, aqueles obtidos junto às imagens com 224x224 pixels não ficaram muito abaixo, podendo ser utilizado em casos em que há limitações de processamento. Com relação à melhor versão do Yolo, para esse conjunto de imagens especificamente, considerou-se o Yolov3 melhor, por apresentar melhores resultados de métricas combinadas. Além disso, foi possível identificar que a CNN conseguiu aprender e extrair as características das placas indicativas com maior facilidade em comparação com as placas de regulamentação e de advertência.

Este trabalho fornece subsídios para a continuação da pesquisa, com investigações de outros tamanhos de imagens; intervalos de iterações superiores a 6000, já que este número é o mínimo recomendado pelo desenvolvedor; além da pesquisa de diferentes distâncias entre as placas e as câmeras, para assim melhor definir a frequência das fotos e assim otimizar as classificações. Realizado esses nossos estudos, a próxima etapa da pesquisa será a análise completa do conjunto de dados das 10.000 imagens rotuladas, com variações na iluminação ao longo do dia.

Vale ressaltar a importância de se ter um modelo como esse sendo utilizado para auxiliar na elaboração dos inventários de sinalização vertical. Um levantamento de inventário utilizando o modo convencional pode demorar meses para ser concluído e que, com a aplicação de métodos de detecção automática, como o apresentado no presente trabalho, pode ser finalizado em horas. Por mais que o modelo não apresente métricas tão próximas de 100%, os resultados com ele obtidos já permitem que se tenha um ganho de tempo e de mão de obra para a execução do trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMBEV S.A.; FALCONI Consultores de Resultados; CLP - Centro de Liderança Pública; TRANSIT Transportes Integrados do DF. Retrato da Segurança Viária – 2017. Brasília, Brasil, 2017.
- A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv, no. May, 2020.
- BRASIL. Lei nº 9503 de 23 de setembro de 1997. Código de Trânsito Brasileiro. Brasília, Brasil, 1997.
- C. R. Ikram, “A benchmark for evaluating Deep Learning based Image Analytics,” UNIVERSITY OF OSLO, 2019.
- Capítulo 40 – INTRODUÇÃO AS REDES NEURAIAS CONVOLUCIONAIS. Deep Learning Book – Data Science. Disponível em: [http://deeplearningbook.com.br/introducao-as-redes-neurais-convolucionais/#:~:text=Uma%20Rede%20Neural%20Convolutacional%20\(ConvNet,de%20diferenciar%20um%20do%20outro.](http://deeplearningbook.com.br/introducao-as-redes-neurais-convolucionais/#:~:text=Uma%20Rede%20Neural%20Convolutacional%20(ConvNet,de%20diferenciar%20um%20do%20outro.) Acesso em: 9 de março de 2021.
- CEARÁ. (2018) Relatório consolidado de segurança viária no estado do Ceará 2014 - 2018. Secretaria da Infraestrutura.
- DENATRAN (2006) CTB – Código de Trânsito Brasileiro: instituído pela Lei n 9.503, de 23 de setembro de 1997. Brasília.
- DNER - Departamento Nacional de Estradas de Rodagem. Diretoria de Desenvolvimento Tecnológico - Divisão de Pesquisas e Desenvolvimento. Guia de redução de acidentes com base em medidas de engenharia de baixo custo. Rio de Janeiro: DCTec, 1998. 140p.
- DNIT – Departamento Nacional de Infraestrutura de Transportes. Manual de Sinalização Rodoviária. Publicação IPR – 743, Ministério dos Transportes, Rio de Janeiro, Brasil, 2010. b.
- FARIA, E.; BRAGA, M. Condições necessárias e objetivos da Educação para o Trânsito desde o ponto de vista dos profissionais brasileiros. Anais do XIII Congresso Pan-americano de Engenharia de Trânsito e Transporte. Albany, NY, 2004.
- FERREIRA, E. R.. Procedimentos automático para apoio na avaliação de pavimentos com o uso de imagens digitais. 2010. Universidade Federal de Viçosa.
- GLERIANI, J.M. Redes neurais artificiais para classificação espectro-temporal de culturas agrícolas. 212p. Tese. Instituto Nacional de Pesquisas Espaciais. São José dos Campos. 2004.
- HOLLERWEGER, M. M.. Aplicação de Visão Computacional no Auxílio ao Levantamento de Defeitos em Pavimento Rodoviário.
- L. A. Silva, H. Sanchez San Blas, D. Peral García, A. Sales Mendes, and G. Villarubia González, “An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images,” Sensors (Basel), vol. 20, no. 21, 2020.
- NODARI, C. Método de Avaliação da Segurança Potencial de Segmentos Rodoviários Rurais de Pista Simples. Tese de D.Sc., Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil, 2003.
- OMS – Organização Mundial da Saúde. Global status report on road safety 2015: Supporting a decade of action. Genebra, Suíça, 2015.
- Pesquisa CNT de rodovias 2019. – Brasília: CNT SEST SENAT, 2019.
- SAMPEDRO, A. T.; CAMPOS, V. B. G. Avaliação e Tratamento das Características da Infraestrutura Viária

- Urbana que Influenciam a Segurança do Tráfego. Engenharia Civil UM (Braga), v. 27, p. 123-134. 2006.
- SANTOS, L. P., SOBRAL, J. L. Inventário Automático de Sinais de Trânsito: um Sistema de Mapeamento Móvel. Universidade do Minho (Braga). 2016
- TAMAYO, A.S. Procedimento para Avaliação e Análise da Segurança de Tráfego em Vias Expressas Urbanas. Tese de Doutorado em Engenharia de Transportes. Universidade Federal do Rio de Janeiro/ COPPE/ Programa de Engenharia de Transportes. Rio de Janeiro, 2010.
- VOINIGESCU, E. Les Neurones ont une Forme Propice à L'apprentissage Profond. 2017. Disponível em: <<https://www.cifar.ca/fr/nouvelles/2017/12/05/les-neurones-ont-une-forme-propice-%c3%a0-1-apprentissage-profond>>.
- Xie, D.; Zhang, L.; Bai, L. Review Article: Deep Learning in Visual Computing and Signal Processing. Hindawi: Applied Computational Intelligence and Soft Computing. Volume 2017, Article ID 1320780. <https://doi.org/10.1155/2017/1320780>
- Mendes, A. S., e González, G. V. (2020) An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images. doi:10.3390/s20216205
- Wang, C., e Liao, H. M. ([s.d.]) YOLOv4: Optimal Speed and Accuracy of Object Detection.
- Zhang, Y., Wang, J., Yang, X., Calculation, O. F., Liu, K., Ye, Y., Yoon, S., e Kim, E. (2018) Understanding of Object Detection Based on CNN Family and YOLO Understanding of Object Detection Based on CNN Family and.

Luciana da Silva Nascimento (luciana.nascimento@alu.ufc.br)

Aline Calheiros Espíndola (aline@det.ufc.br)

Ricardo Venescau de Oliveira Almeida (venescau@det.ufc.br)

Ernesto Ferreira Nobre Júnior (nobre@ufc.br)

Departamento de Engenharia de Transportes, Universidade Federal do Ceará

Campus do Pici, Bloco 703 – Fortaleza, CE, Brasil