

Detecção Inteligente de Falhas em Pavimentações Asfálticas com Redes Neurais Convolucionais Regionais

Rafael Barbosa de Carvalho, Elloá B. Guedes, Carlos Maurício S. Figueiredo

¹Grupo de Pesquisa em Sistemas Inteligentes
Universidade do Estado do Amazonas (UEA)
Av. Darcy Vargas, 1200 – Manaus – Amazonas
{rbc.eng, ebgcosta, cfigueiredo}@uea.edu.br

Abstract. *In this paper we addressed the automatic road damage inspection as a Computer Vision detection problem in benefit of solutions to help smart cities improve traffic quality and security. To do so, we considered an experimental scenario with realistic data from three different countries and four configurations of YOLO networks. When compared to related work from literature, our results have significant improvements in prediction time using a lower number of parameters, yielding an experimental mAP of 0.53. We also evaluated our solution in a case study with images from Brazil that highlights several practical challenges that need to be taken into account when proposing automatic detection models for such problem.*

Resumo. *Neste artigo foi abordado o problema da inspeção automática de danos em pavimentações como um problema de detecção em Visão Computacional com vistas a colaborar para o desenvolvimento de soluções que apoiem cidades inteligentes na melhoria da qualidade e da segurança no trânsito. Neste sentido considerou-se um cenário experimental com dados realísticos oriundos de três países diferentes e quatro configurações para as redes YOLO. Ao comparar os resultados obtidos com a literatura, verificam-se ganhos significativos no tempo de previsão ainda que utilizando um menor número de parâmetros, produzindo um mAP igual a 0,53. Também avaliamos a solução proposta em um estudo de caso com imagens do Brasil, o qual ressaltou diversos desafios práticos a serem levados em conta na ocasião da proposição de modelos automáticos de detecção para o problema em consideração.*

1. Introdução

Segundo a Organização das Nações Unidas (ONU), atualmente 55% da população mundial vive em áreas urbanas e projeta-se que no ano de 2050 esta proporção seja de 68% [ONU 2018]. Diante desta perspectiva, é mister que se faça uso dos desenvolvimentos e avanços das tecnologias da Informação e Comunicação para que os serviços providos nas áreas urbanas tradicionais se tornem mais flexíveis, eficientes, sustentáveis e inteligentes, como preconizado para as *Cidades Inteligentes* (CI), nas quais há a melhoria contínua em áreas tais como infraestrutura, transporte, energia e assistência médica, trazendo benefícios aos seus habitantes [Janani et al. 2021, Pellicer et al. 2013].

Particularmente, na área de transporte, uma CI busca contemplar melhorias no acesso, eficiência de locomoção, segurança e conforto para os usuários. Quando os meios de transporte são percebidos pela população itinerante como uma fonte de incômodo em decorrência da degradação avançada, da inconveniência para mobilidade e da infraestrutura insuficiente,

umentam-se os desafios e, conseqüentemente, a demanda por soluções viáveis. Nesse contexto, as condições precárias das pavimentações asfálticas e os defeitos estruturais em rodovias, por exemplo, reduzem não apenas a segurança e o conforto, como também exercem um impacto negativo na operação de veículos automotores [Staniek 2021].

Em relação à pavimentação asfáltica, é comum que haja desgaste e deterioração devido a fatores relacionados à localização, condições climáticas, volume de tráfego, soluções de Engenharia, dentre outros. Tal fato demanda uma extensa análise de qualidade para uma manutenção eficiente e econômica visando preservar seu estado e, por conseguinte, a segurança dos que nela transitam [Miller and Zaloshnja 2009]. De maneira geral, muitos municípios e autoridades procuram implementar avaliações automatizadas dos danos nas pavimentações. No entanto, geralmente se deparam com as limitações de tecnologia, de *know-how* e de recursos financeiros para adquirir equipamentos de última geração para coleta e análise de dados em tais contextos [Arya et al. 2021a].

Os desafios da análise de qualidade e manutenção da pavimentação asfáltica são predominantes no Brasil, no qual o transporte rodoviário é o principal sistema logístico do País, correspondendo a cerca de 65 % de tudo o que circula [CNT 2019]. A malha rodoviária federal é constituída por 120.767,3 km, dos quais apenas apenas 64.000 km (52,9 %) são pavimentados. Quando a pavimentação é deficiente, aumentam os custos de manutenção dos veículos, o consumo de combustíveis, o uso de lubrificantes, o desgaste de pneus e freios e ainda há redução da segurança viária. No ano de 2020, por exemplo, registrou-se 63.447 acidentes nas rodovias federais brasileiras, que culminaram na perda irreparável de vidas e em custos financeiros da ordem de R\$ 10,22 bilhões [CNT 2020].

Desta feita, para colaborar na resolução do problema da inspeção automática da qualidade de pavimentações asfálticas, o objetivo deste trabalho consiste em avaliar o desempenho de Redes Neurais Convolucionais Regionais (R-CNNs, do inglês *Regional Convolutional Neural Networks*) da família YOLO (acrônimo para *You Only Look Once*) na tarefa de Visão Computacional de detecção (localização e classificação) de quatro tipos distintos de falhas. Nesta solução, considerou-se a aquisição de experiência a partir de 18.667 exemplos oriundos da base de dados RDD2020 [Arya et al. 2021b]. Os resultados obtidos ressaltaram a arquitetura YOLOv5 como tendo melhor desempenho na tarefa proposta, com *mean Average Precision* (mAP) igual a 53,19 % no cenário experimental de avaliação, ressaltando o potencial de utilização em contextos práticos, inclusive em dispositivos embarcados em decorrência do seu baixo número de parâmetros. Além disso, foi efetuada uma validação preliminar do modelo proposto com imagens de pavimentações asfálticas coletadas no Brasil com vistas a aferir as vantagens e limitações da generalização em um contexto prático em que a demanda por tais soluções é mister.

Para apresentar o que se propõe, este artigo está organizado como segue. Trabalhos afins na literatura são apresentados e discutidos na Seção 2. A metodologia que guiou a realização deste trabalho, contemplando os dados experimentais, os modelos e as práticas para avaliação de desempenho, encontra-se descrita na Seção 3. Os resultados são apresentados e discutidos na Seção 4. O estudo de caso da avaliação de pavimentações asfálticas no Brasil é detalhado na Seção 5. Por fim, as considerações finais são mostradas na Seção 6.

2. Trabalhos Relacionados

O problema considerado no escopo deste trabalho já foi abordado na literatura com diferentes técnicas. Conforme o *survey* de Cao *et al.* [2020], as soluções seminais consideravam o uso

de sensores, tais como acelerômetro e giroscópio, para reconhecer danos na pavimentação por meio da amplitude da vibração detectada. Em seguida, com o avanço das técnicas de processamento de imagem, predominaram as soluções tradicionais de Visão Computacional baseadas nos métodos de extração de características em conjunto com as técnicas de *Machine Learning* para classificação. Com o advento de *Deep Learning*, houve melhoria significativa no desempenho das soluções inteligentes para o problema, mas as contribuições variavam quanto aos métodos (R-CNNs, SSDs, R-FCNs), estratégias de validação e dados experimentais, o que dificultava uma comparação mais objetiva entre as mesmas.

Visando a expansão global das soluções de Visão Computacional para inspeção de pavimentações asfálticas, o IEEE promoveu uma competição internacional denominada *Global Road Damage Detection Challenge* (GRDDC) [Maeda et al. 2020] na qual imagens capturadas em diferentes localidades com *smartphones* e câmeras veiculares estavam à disposição para elaboração de métodos automáticos de detecção de falhas em tais pavimentações. A competição contou com a participação de 121 equipes, cujas 12 melhores soluções foram sintetizadas em uma publicação recente [Arya et al. 2020], a qual será tomada como referência no decorrer desta seção para reportar o estado da arte das soluções para o problema dada a uniformidade nos quesitos de dados experimentais e de avaliação das soluções.

As equipes tiveram acesso à base de dados RDD2020 [Arya et al. 2021b] para elaboração de suas soluções, treinamento e validação dos modelos. A etapa de testes foi realizada em duas bases de dados privadas, com 2.631 e 2.664 exemplos cada, e o ranqueamento das equipes deu-se pela aferição do F_1 -Score das previsões em comparação com os valores observados nos rótulos destes conjuntos de testes. Ao examinar as melhores soluções elencadas, 67% das equipes utilizaram modelos da família YOLO. Além disso, as 3 melhores soluções fizeram uso de *ensemble learning* gerando comitês de R-CNNs e explorando múltiplas estratégias para aumento artificial de dados, o que colabora na regularização do treinamento dos modelos e, assim, favorece a generalização perante novos exemplos.

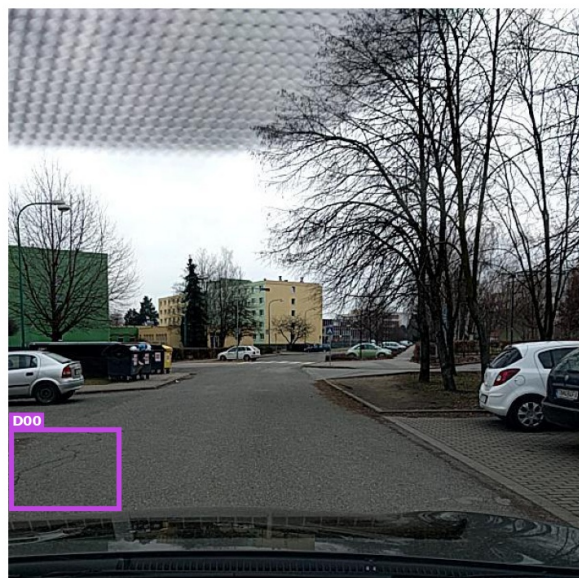
A equipe vencedora, em particular, composta por integrantes de universidades dos Estados Unidos e da Jordânia, utilizou duas estratégias principais, a citar: (1) *Test Time Augmentation*, na qual, para cada imagem de teste, geram-se múltiplas previsões que são filtradas com um algoritmo de supressão não-máxima para produzir a melhor previsão final; (2) comitês de modelos, nos quais a mesma entrada é fornecida a diferentes configurações de redes YOLOv5 e as previsões resultantes são combinadas por meio de uma média, o que auxilia a obter uma melhor acurácia, reduzindo a variância das previsões. Como resultado, a equipe obteve F_1 -Score aproximadamente igual a 0,67 em ambas as partições de testes [Hegde et al. 2020]. Apesar de ter obtido os melhores indicadores de desempenho, reportou-se uma explícita queda no tempo das previsões em comparação às outras equipes e não foram aferidos aspectos relativos ao número de parâmetros da solução final proposta [Arya et al. 2020].

3. Materiais e Métodos

O problema da inspeção automática da qualidade de pavimentações asfálticas foi abordado no escopo deste trabalho como uma tarefa de detecção (localização e classificação) mediante Aprendizado Supervisionado com R-CNNs da família YOLO, modelos do estado da arte de *Deep Learning* para o referido tipo de tarefa. Os dados experimentais, modelos e a avaliação de desempenho considerados são descritos detalhadamente nas subseções a seguir.

3.1. Dados Experimentais

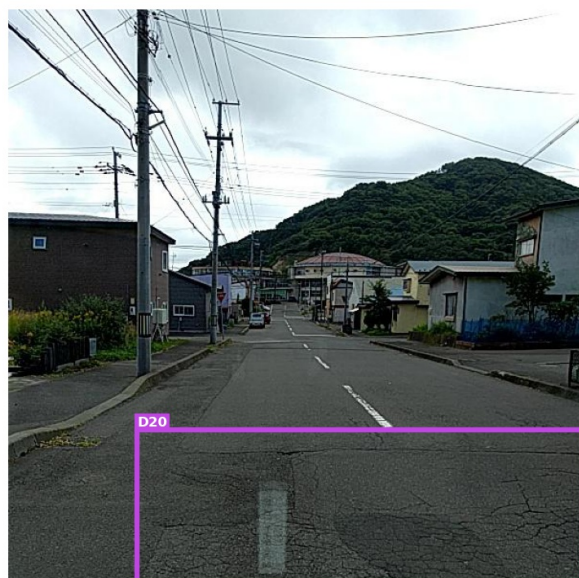
Como fonte de experiência para a tarefa foi utilizado o *dataset* RDD2020 [Arya et al. 2021b], o qual possui originalmente 26.336 imagens de pavimentações de três diferentes países, a citar: República Tcheca, Índia e China. Neste *dataset* há mais de 31.000 instâncias de falhas de quatro classes, as quais são ilustradas na Figura 1 e são descritas como segue:



(a) D00



(b) D10



(c) D20



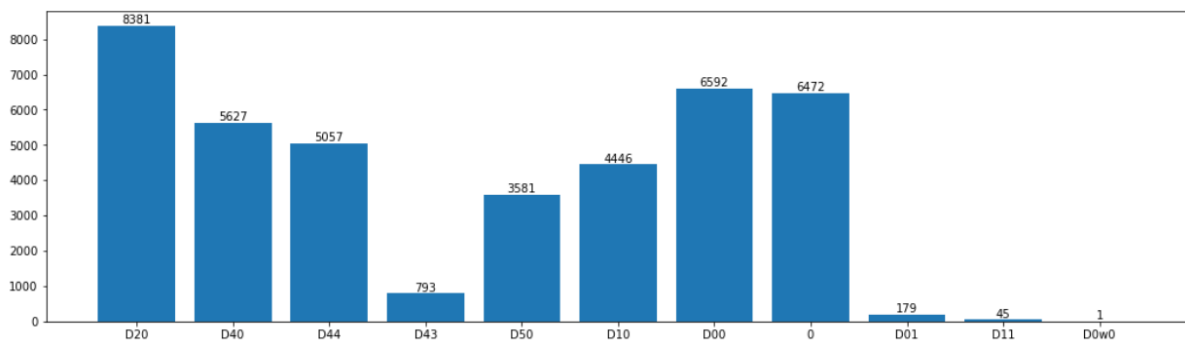
(d) D40

Figura 1: Exemplos de imagens da base de dados RD2020 ilustrando as quatro classes a serem detectadas.

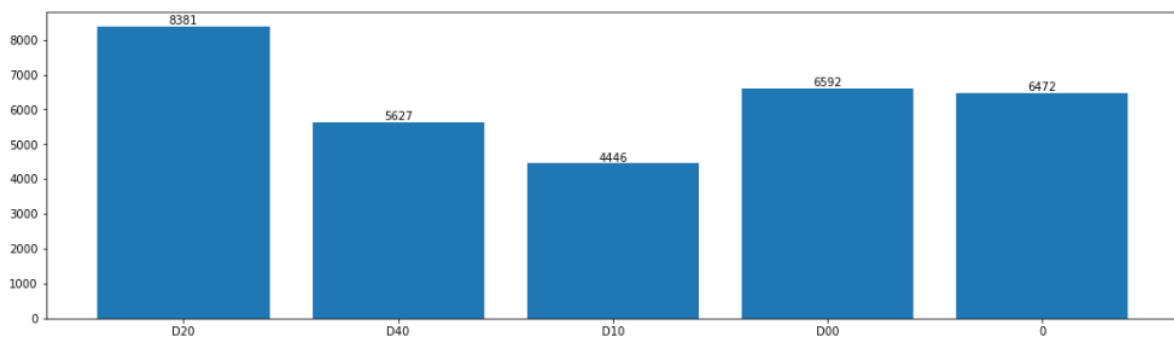
1. **Classe D00.** Rótulo associado a trincas longitudinais (*longitudinal cracks*), as quais são predominantemente paralelas ao eixo da via, sendo consideradas um dano funcional e estrutural no pavimento;

2. **Classe D10.** Rótulo associado a trincas transversais (*transverse cracks*), as quais são perpendiculares ao eixo da via e causam danos funcionais e estruturais no pavimento;
3. **Classe D20.** Rótulo associado a trincas em malha tipo “couro de jacaré” (*alligator cracks*), as quais são múltiplas, interligadas e sem direções definidas. Representam um dano estrutural tipicamente característico do fim da vida útil do asfalto;
4. **Classe D40.** Rótulo associado a buracos (*potholes*), isto é, cavidades de tamanhos variados no revestimento do pavimento [CNT 2018].

Após a obtenção dos exemplos da base de dados, foi realizada uma etapa de pré-processamento e análise exploratória da mesma, visando identificar a quantidade de classes e sua distribuição, cujos resultados são ilustrados no histograma da Figura 2a. Em seguida, foram consideradas apenas as imagens com rótulos disponíveis, contabilizando 21.041 itens. Desse conjunto, foram desconsideradas ocorrências de classes não pertencentes ao conjunto indicado anteriormente (D00, D10, D20 e D40) por não ser possível determinar seu impacto nos resultados subsequentes devido à falta de informação sobre as mesmas. Além disso, foi adicionada uma nova classe, rotulada como “0”, que indica que a respectiva imagem não possui instâncias de falhas na pavimentação. Ao final deste pré-processamento, conforme ilustra o histograma da Fig. 2b, tem-se 18.667 imagens com 25.046 ocorrências de falhas. Nota-se que as classes encontram-se desbalanceadas, o que impõe uma dificuldade prática na detecção daquelas com menor quantidade de exemplos.



(a) Base de dados RD2020.



(b) Dados experimentais após limpeza e pré-processamento

Figura 2: Histograma da distribuição de exemplos por classe.

3.2. Modelos, Parâmetros e Hiperparâmetros

A proposta *You Only Look Once* (YOLO) foi desenvolvida conforme uma abordagem para R-CNNs na qual o modelo deve realizar todas as tarefas necessárias (detecção e classificação

de múltiplos objetos) em uma única etapa (*single shot*). Para tanto, aborda a detecção de objetos como um problema de regressão, utilizando uma única rede neural para prever as *bounding boxes* e as classes dos objetos [Redmon et al. 2016]. Em decorrência da eficiência dessa estratégia, a classificação é rápida e tem favorecido a utilização em diversas aplicações de tempo real, superando o desempenho das arquiteturas previamente propostas nesse nicho [Michelucci 2019].

A YOLO foi sujeita a sucessivas melhorias desde sua proposição. Na YOLOv4, por exemplo, houve enfoque em melhorias que promovessem maior velocidade de inferência e aumento na acurácia. Para tanto, foi projetada para treino exclusivamente com aceleração em *hardware* via GPU, considerou uma nova abordagem de aumento artificial de dados e também novas estratégias para extração de características e regularização [Bochkovskiy et al. 2020]. A YOLOv5, por sua vez, foi projetada para dar prosseguimento às melhorias na YOLOv4, especialmente com vistas a evitar conflitos com as constantes atualizações de *frameworks*. Esta proposta contemplou melhorias em relação à acurácia, velocidade de treinamento e de inferência, além de uma redução na quantidade de pesos. De uma perspectiva tecnológica, foi desenvolvida nativamente com o *Pytorch*, visando maior suporte e um processo de *deploy* simplificado [Jocher et al. 2020].

Por constituírem o estado da arte para detecção de objetos conforme a abordagem YOLO, as versões v4 e v5 foram os modelos considerados no escopo deste trabalho. Em ambas as arquiteturas, foram utilizados pesos pré-treinados na base de dados MS COCO. A referida estratégia de transferência de aprendizado é amplamente sugerida para melhorar o desempenho da detecção em um domínio alvo por aproveitar características previamente aprendidas pela arquitetura em uma base de dados de origem.

A técnica *Early Stopping* foi adotada para regularizar o treinamento do modelo e considerou o monitoramento da métrica mAP no conjunto de validação ao longo das épocas. No tocante aos hiperparâmetros, considerou-se quatro configurações diferentes, as quais encontram-se listadas a seguir:

- **Arquitetura: YOLOv4:**

- **Configuração 1.** Tamanho do *batch*: 64; Tamanho máximo do *batch*: 4.000; Taxa de aprendizado: 10^{-3} ; *Steps*: 6.400;
- **Configuração 2.** Tamanho do *batch*: 64; Tamanho máximo do *batch*: 4.000; Taxa de aprendizado: 10^{-2} ; *Steps*: 3.200;

- **Arquitetura: YOLOv5:**

- **Configuração 3.** Tamanho da arquitetura: Pequeno; Tamanho do *batch*: 16; Taxa de aprendizado: 10^{-2} ; *Épocas*: 300;
- **Configuração 4.** Tamanho da arquitetura: Médio; Tamanho do *batch*: 16; Taxa de aprendizado: 10^{-2} ; *Épocas*: 300;

3.3. Avaliação de Desempenho

Para avaliação de desempenho, foi utilizada a técnica de validação cruzada *holdout*, com 70% dos dados destinados ao treinamento, 10% para validação e 20% para testes. As métricas de desempenho foram aferidas na partição de testes e contemplam: Precisão, Revocação, F_1 -Score e mAP com *threshold* $t \geq 0.5$ para o IoU, a qual será denotada como mAP@0.5. Uma explicação detalhada do cálculo e semântica de tais métricas no contexto da detecção de objetos em Visão Computacional encontra-se disponível no *survey* de Padilla *et al.* [2020].

4. Resultados e Discussão

Para a execução dos *scripts* de treinamento foi utilizado um servidor com a seguinte configuração: processador Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz, 32 GB de memória principal, 960 GB de memória secundária e 2 placas de vídeo NVIDIA GTX 1080 Ti com 11 GB de VRAM para aceleração em *hardware* do treinamento. Os resultados obtidos do treinamento para as diferentes configurações propostas encontram-se dispostos na Tabela 1.

Tabela 1: Síntese dos resultados experimentais.

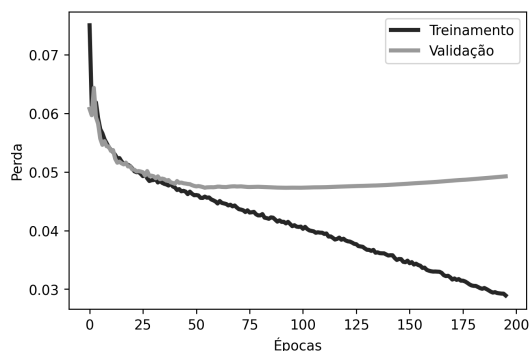
Métrica	YOLOv4		YOLOv5	
	Configuração 1	Configuração 2	Configuração 3	Configuração 4
Precisão	66,00 %	59,00 %	54,11 %	57,21 %
Revocação	18,00 %	22,00 %	52,28 %	52,08 %
F₁-Score	28,00 %	32,00 %	53,18 %	54,53 %
mAP@0.5	31,27 %	31,43 %	50,70 %	53,19 %

Ao analisar tais resultados considerando a métrica de desempenho elencada como referência, percebe-se que a melhor arquitetura para a tarefa em questão foi a YOLOv5 com Configuração 4, em que o mAP@0.5 no conjunto de testes foi 4,91 % e 70,09 % superiores aos desempenhos da segunda e última colocadas, respectivamente. Ao examinar os artefatos visuais derivados dos experimentos com esta rede, observa-se que houve preempção nas épocas de treinamento para evitar *overfitting*, vide Figs. 3a e 3b e que, conforme matriz de confusão da Figura 3c, as falhas do tipo trincas em malha (Classe D20) foram as melhores detectadas na etapa de testes, mas que também houve falsos negativos na detecção, indicados pelo rótulo BFN (*Background False Negative*). A partir dos experimentos, também foi possível identificar que esta arquitetura pode realizar predições a uma taxa de 64 *frames* por segundo (FPS).

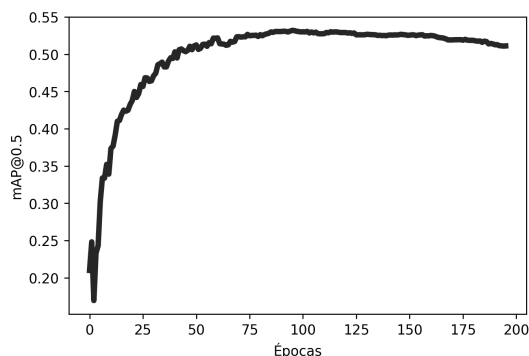
O melhor desempenho da YOLOv5, perante a Configuração 4 no cenário experimental, pode ser justificado em razão de uma melhor eficiência dos procedimentos para aumento artificial de dados quando comparada à YOLOv4 e quanto ao número de parâmetros quando comparada à Configuração 3. Ao analisar os resultados em comparação com o estado da arte para o RDD2020 [Hegde et al. 2020], observa-se que as métricas de desempenho ficam aquém da melhor solução na literatura, com decréscimo percentual de 18,61 % no *F₁-Score*. Porém, há de se ressaltar que (i) a solução proposta contorna as limitações no tocante ao tempo de previsão, com 64 FPS versus 3,08 FPS (aumento percentual de 1.977,92 %); (ii) possui significativamente menos parâmetros, uma vez que a solução de referência usa um comitê de 3 redes YOLOv5; e (iii) que foi treinada com significativamente menos dados, pois houve limpeza dos exemplos com rótulos inválidos, vide comparativo da Fig. 2, e que, para contornar o fato das bases de dados de teste do GRDDC serem privadas, foi necessário separar uma partição para teste a partir dos dados de treinamento. Neste último ponto, o trabalho de referência da literatura foi exposto a 30 % mais exemplos de treinamento do que a solução aqui proposta, aspecto crítico para o desempenho quando se tem modelos baseados em Aprendizado Supervisionado.

5. Um Estudo de Caso em Pavimentações Asfálticas no Brasil

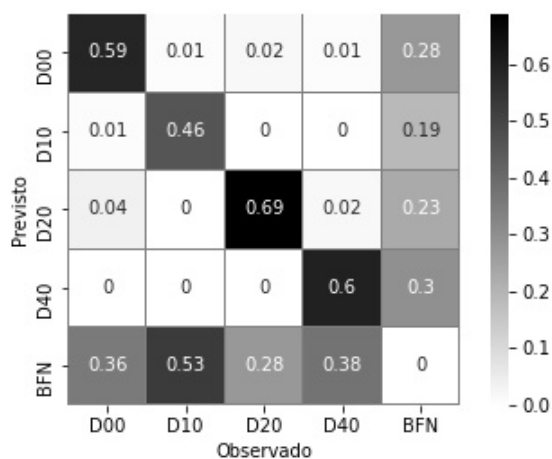
Para avaliar o desempenho da melhor solução identificada anteriormente no contexto de pavimentações asfálticas brasileiras, foi utilizada a base de dados *Road Traversing Knowledge* (RTK *dataset*) [Rateke et al. 2019]. As imagens presentes nessa base de dados foram coletadas



(a) Gráfico da Perda no Treinamento



(b) Gráfico do mAP@0.5 no Treinamento



(c) Matriz de Confusão

Figura 3: Artefatos do treinamento e teste da YOLOv5 com Configuração 4.

em duas cidades vizinhas à Florianópolis, SC com o uso de uma *webcam*, durante o dia e em boas condições climáticas. Inicialmente proposta para fins de segmentação quanto ao tipo de pavimento e composta por imagens com dimensões 352×288 *pixels*, essa base de dados foi pré-processada da seguinte forma: (1) foram mantidos os exemplos apenas de pavimentações asfálticas; (2) os rótulos de segmentação foram convertidos para *bounding boxes*; (3) houve um mapeamento das classes do RTK para as classes do RDD2020; e (4) exemplos de outras classes foram descartados. Em particular, o mapeamento de classes foi realizado como segue: os exemplos rotulados como *potholes* foram mapeados para a Classe D40; os exemplos rotulados como *cracks* foram mapeados para a Classe D00 quando a largura era maior que o comprimento e para a Classe D10 em caso contrário. Uma descrição estatística da preparação do RKT *dataset* em termos do quantitativo das classes e do tamanho das *bounding boxes* para o estudo de caso encontra-se disposta na Tabela 2. A Fig. 4 contempla alguns exemplos presentes na base de dados com os rótulos desejados.

Os exemplos resultantes da preparação do RKT *dataset* foram sujeitos à uma partição *holdout* com as mesmas proporções descritas na Seção 3.3. Os exemplos de treino foram utilizados então para promover uma Transferência de Aprendizagem na solução de referência resultante da Configuração 4. Em seguida, foi efetuada uma avaliação de desempenho, cu-



Figura 4: Exemplos do RTK *dataset* com rótulos de classes mapeados.

Tabela 2: Descrição estatística do pré-processamento do RTK *dataset*.

	Comprimento			Largura			Exemplos
	Média	Máx	Mín	Média	Máx	Mín	Quantidade
Classe D00	43.27 ± 26.06	98	6	33.61 ± 21.17	87	3	36
Classe D10	18.76 ± 21.00	107	2	42.24 ± 47.70	345	5	390
Classe D40	13.92 ± 10.11	62	3	34.52 ± 17.38	93	4	105

jas métricas encontram-se sintetizadas na Tabela 3. Exemplos de previsões produzidas pelo modelo para o cenário do estudo de caso encontram-se ilustradas na Fig. 5.

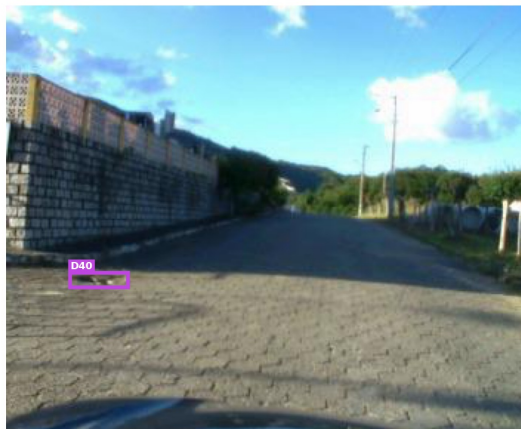
Tabela 3: Desempenho da rede na previsão de falhas em pavimentações asfálticas no Brasil.

	Precisão	Revocação	F ₁ -Score	mAP@0.5
Todas as Classes	43,8 %	51,0 %	47,1 %	45,6 %
Classe D00	18,7 %	23,4 %	20,8 %	11,5 %
Classe D10	42,6 %	45,0 %	43,8 %	40,9 %
Classe D40	70,2 %	84,6 %	76,7 %	84,3 %

Ao observar os resultados obtidos neste estudo de caso, é possível observar que o mAP@0.5 sofreu um decréscimo percentual de 14,25% quando comparado aos resultados apresentados anteriormente, ou seja, a transferência de aprendizado foi negativa. Alguns fatores podem justificar este comportamento, tais como a baixa resolução das imagens; o tamanho das *bounding boxes*; e, mais predominantemente, a diferença entre os exemplos do RDD2020 no sentido de prover experiência representativa para favorecer a generalização no cenário brasileiro. Tais achados motivam a elaboração de bases de dados representativas para o Brasil no tocante às falhas em pavimentações asfálticas e ressaltam as dificuldades em construir bases de dados de abrangência global para este problema.

6. Considerações Finais

O trabalho em questão apresentou os resultados experimentais obtidos do treinamento e avaliação das R-CNNs YOLOv4 e YOLOv5 para detecção automática de falhas em pavimentações asfálticas. Para tanto, foi considerada uma base de dados com mais de 18 mil imagens e quatro tipos de defeitos (trincas longitudinais, transversais, em malha e buracos). Foi possível verificar que a arquitetura YOLOv5 obteve melhores resultados nesta tarefa,



(a) Exemplo 1 – Desejado



(b) Exemplo 1 – Previsto



(c) Exemplo 2 – Desejado



(d) Exemplo 2 – Previsto

Figura 5: Exemplos de previsão do modelo no cenário brasileiro.

com mAP igual a 0.53 para o conjunto de testes, o que pode ser decorrente das melhorias desse modelo em relação ao seu predecessor e também do número de parâmetros existentes. Um estudo de caso também foi conduzido com vistas a aferir o potencial de uso do modelo de detecção proposto em um contexto prático do Brasil, o qual resultou em desempenho aquém do desejado, o que pode ter ocorrido em virtude da baixa resolução das imagens e da distância entre domínio de origem e domínio de aplicação.

Em trabalhos futuros, almeja-se melhorar o desempenho da solução, buscando explorar mais técnicas de aumento de dados, configurações e ajuste fino de parâmetros. No contexto brasileiro, pretende-se explorar técnicas de Transferência de Aprendizado a partir de pesos obtidos em outros cenários e também ressalta-se a necessidade de propor bases de dados com maior quantitativo de imagens para o problema, com vistas a construir soluções inteligentes para fomentar a inspeção automática da qualidade das pavimentações no País.

Agradecimentos

Os autores agradecem ao Laboratório de Sistemas Inteligentes da Universidade do Estado do Amazonas pela disponibilização dos recursos computacionais que viabilizaram a realização deste trabalho.

Referências

- Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T., and Sekimoto, Y. (2021a). Deep learning-based road damage detection and classification for multiple countries. *Automation in Construction*, 132:103935.
- Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Omata, H., Kashiyama, T., and Sekimoto, Y. (2020). Global road damage detection: State-of-the-art solutions. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 5533–5539. IEEE.
- Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., and Sekimoto, Y. (2021b). RDD2020: An annotated image dataset for automatic road damage detection using deep learning. *Data in Brief*, 36:107133.
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
- Cao, M.-T., Tran, Q.-V., Nguyen, N.-M., and Chang, K.-T. (2020). Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources. *Adv. Eng. Inform.*, 46(101182):101182.
- CNT (2018). Confederação Nacional do Transporte (cnt) – conheça os 13 principais defeitos do pavimento das rodovias. Disponível em <https://www.cnt.org.br/agencia-cnt/conheca-principais-defeitos-pavimento>. Acesso em 19 de maio de 2022.
- CNT (2019). Confederação Nacional do Transporte (cnt – piora a qualidade das rodovias brasileiras. Disponível em <https://www.cnt.org.br/agencia-cnt/piora-a-qualidade-das-rodovias-brasileiras>. Acesso em 19 de maio de 2022.
- CNT (2020). Confederação Nacional do Transporte (cnt) – CNT lança painel com dados do transporte rodoviário no Brasil. Disponível em <https://cnt.org.br/agencia-cnt/cnt-lanca-painel-com-dados-do-transporte-rodoviario-no-brasil>. Acesso em 19 de maio de 2022.
- Hegde, V., Trivedi, D., Alfarrarjeh, A., Deepak, A., Ho Kim, S., and Shahabi, C. (2020). Yet another deep learning approach for road damage detection using ensemble learning. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE.
- Janani, R. P., Renuka, K., Aruna, A., and Lakshmi, N. K. (2021). IoT in smart cities: A contemporary survey. *Global Transitions Proceedings*. Disponível em <https://doi.org/10.1016/j.glt.2021.08.069>. Acesso em 19 de maio de 2022.
- Jocher, G., Stoken, A., Borovec, J., NanoCode012, ChristopherSTAN, Changyu, L., Laughing, tkianai, Hogan, A., lorenzomamma, yxNONG, AlexWang1900, Diaconu, L., Marc, wanghaoyang0106, ml5ah, Doug, Ingham, F., Frederik, Guilhen, Hatovix, Poznanski, J., Fang, J., Yu, L., changyu98, Wang, M., Gupta, N., Akhtar, O., PetrDvoracek, and Rai, P. (2020). ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements. Disponível em <https://doi.org/10.5281/zenodo.4154370>. Acesso em 19 de maio de 2022.
- Maeda, H., Arya, D., Sekimoto, Y., Ghosh, S. K., Toshniwal, D., Seto, T., Kashiyama, T., and Omata, H. (2020). Global road damage detection challenge 2020. IEEE BigData 2020. Disponível em <https://rdd2020.sekilab.global/>. Acesso em 19 de maio de 2022.

- Michelucci, U. (2019). *Advanced applied deep learning: convolutional neural networks and object detection*. Springer.
- Miller, T. R. and Zaloshnja, E. (2009). On a crash course: The dangers and health costs of deficient roadways. Technical Report 01135610. Disponível em <http://www.infrastructureusa.org/on-a-crash-course-the-dangers-and-health-costs-of-deficient-roadways/>. Acesso em 19 de maio de 2022.
- ONU (2018). *The World's Cities in 2018*, volume 1. Department of Economical and Social Affairs – Population Dynamics. Disponível em <https://population.un.org/wup/Publications/>. Acesso em 19 de maio de 2022.
- Padilla, R., Netto, S. L., and da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 237–242, Niterói, Brasil.
- Pellicer, S., Santa, G., Bleda, A. L., Maestre, R., Jara, A. J., and Skarmeta, A. G. (2013). A global perspective of smart cities: A survey. In *2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pages 439–444. Disponível em <https://10.1109/IMIS.2013.79>. Acesso em 19 de maio de 2022.
- Rateke, T., Justen, K. A., and Von Wangenheim, A. (2019). Road surface classification with images captured from low-cost camera - Road Traversing Knowledge (RTK) dataset. *Rev. Inform. Teór. Apl.*, 26(3):50–64.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Estados Unidos. IEEE.
- Staniek, M. (2021). Road pavement condition diagnostics using smartphone-based data crowdsourcing in smart cities. *J. Traffic Transp. Eng. (Engl. Ed.)*, 8(4):554–567.