OpenImages Cyclists: Expandindo a Generalização na Detecção de Ciclistas em Câmeras de Segurança

Ednilza Evangelista da Silva Nardi[1], **Bruno Padilha**[1], **Leonardo Tadashi Kamaura**[1], **João Eduardo Ferreira**[1]

[1]Departamento de Ciência da Computação – IME – USP
São Paulo – SP –Brasil

{ednilza,jef}@ime.usp.br, brunopadilha@usp.br, ltkamaura@alumni.usp.br

Abstract. *Although there are several public datasets containing cyclists for training detectors based on Deep Learning, their notes are for bicycles and people, or the quality and quantity of images are limited. To overcome these limitations, we propose the new dataset OpenImages Cyclists, built by a preselection of images from the OpenImages dataset and a new algorithm for the semi-automated generation of cyclist notes assisted by people and bicycle detectors. By training a detector with this data, we got an identification rate of 78% for the detection of cyclists in the São Paulo - Capital campus at USP through learning transfer, higher than 52% reached with the Mio-TCD datasets.*

Resumo. *Embora haja diversos conjuntos de dados públicos contendo ciclistas para treinamento de detectores baseados em Aprendizado Profundo, suas anotações são para bicicletas e pessoas, ou a qualidade e quantidade das imagens são limitadas. Para superar essas limitações, propomos o novo conjunto de dados OpenImages Cyclists, construído por meio de pré-seleção de imagens do conjunto OpenImages e de um novo algoritmo para geração semi-automatizada de anotações de ciclistas auxiliado por detectores de pessoas e bicicletas. Ao treinar um detector com esses dados, obtivemos uma taxa de identificação da ordem de 78% na detecção de ciclistas na USP, Campus São Paulo - Capital, por transferência de aprendizado, maior que os 52%, com o conjunto MIO-TCD.*

1. Introdução

No trânsito, uma boa detecção automatizada de ciclistas permite aumentar a segurança tanto do próprio ciclista quanto de pedestres e veículos. Trata-se de um usuário vulnerável de vias públicas [Li et al. 2016] e, assim como os pedestres, está sujeito a situações de risco, porém com velocidade e ocupação de espaço diferentes [Masalov et al. 2019]. Além disso, no contexto do monitoramento de vias por câmeras de segurança, existem regras próprias para circulação de ciclistas, diferentes das regras para pedestres e para veículos motorizados.

A detecção automatizada de objetos tem sido muito estudada em diversos cenários como, por exemplo, monitoramento automatizado de vias públicas, por câmeras de segurança, e condução autônoma, principalmente com métodos de aprendizado profundo (do inglês *Deep Learning*) [Santhosh et al. 2020]. Essa técnica é uma das que mais tem tido sucesso na tarefa de detecção de objetos, principalmente por sua propriedade de generalização [Zhang et al. 2021]. Entretanto, a qualidade da detecção nos modelos de

aprendizado profundo depende de grande quantidade, qualidade e variabilidade dos dados de treinamento [Zhou et al. 2017].

Existe uma grande quantidade de dados em bases publicamente acessíveis para treinar modelos de aprendizado profundo que reconhecem e localizam pessoas, bicicletas, veículos além de diversos outros objetos. Por exemplo *VOC-Pascal* [Everingham et al. 2010], *COCO* [Lin et al. 2014] e *Open Images* [Kuznetsova et al. 2020]. Já no caso de ciclista, existem poucos dados disponíveis e os poucos que existem, como *Tsinghua-Daimler* [Li et al. 2016] e *Specialized Cyclist* [Masalov et al. 2019] foram coletados com uma câmera montada na frente de um veículo em regiões geográficas restritas, os quais registram os ciclistas em um ângulo muito distinto do comumente encontrado em câmeras de segurança (i.e. em postes altos e apontando para baixo). Isso, conforme demonstrado na seção 4, afeta negativamente a capacidade de generalização para ambientes de câmeras de segurança. Já o conjunto *MIO-TCD* [Luo et al. 2018], embora proveniente de câmeras de segurança, contém pouco menos de duas mil imagens de bicicletas incluindo a pessoa condutora na mesma anotação. Essas imagens são de baixa resolução, baixa qualidade, uma vez que contém muitos artefatos de compressão de video, além dos ciclistas aparecem, em geral, muito pequenos e com poucos detalhes.

Treinar um detector de objetos capaz de localizar com boa acurácia pessoas e bicicletas pode não ser suficiente para a correta identificação de ciclistas. O ciclista é um objeto composto resultante de uma interação especifica de uma pessoa conduzindo uma bicicleta. Por exemplo, uma pessoa parada ao lado de uma bicicleta não é um ciclista, assim como uma pessoa empurrando uma bicicleta na calçada. Desse modo, faz-se necessário que o detector de ciclistas aprenda, além das características de uma bicicleta, as características de uma pessoa na posição especifica de conduzir a bicicleta, possivelmente vestindo trajes e equipamentos de proteção adequados a essa prática esportiva.

O conjunto de dados *Open Images* disponibiliza aproximadamente 18.000 imagens contendo ciclistas, das quais cerca de um terço possui todas as bicicletas apresentadas em cada imagem devidamente anotadas, ou seja, caixas delimitadoras (do inglês *bounding-boxes*) sem cortar partes dos objetos e com tamanho mínimo necessário. Para essas mesmas imagens, poucas pessoas estavam anotadas ou com caixas delimitadoras ruins. É muito importante que a maioria dos objetos, alvos em uma mesma imagem, estejam anotados para que um detector baseado em aprendizado profundo aprenda satisfatoriamente a identificá-los. Alternativamente, a detecção de pessoas pode ser feita com modelos treinados no conjunto de dados *COCO*. Por exemplo, o detector de objetos *YOLOv4* [Bochkovskiy et al. 2020] disponibiliza uma configuração de pesos pré-treinados exaustivamente no *COCO*, dispensando assim a necessidade de anotar manualmente as pessoas nas imagens contendo ciclistas do *Open Images*.

Desse modo, este trabalho apresenta um novo algoritmo para a geração semi-automatizada de anotações de ciclistas presentes em um subconjunto de imagens públicas disponíveis no *Open Images*. Por meio de uma pré-seleção manual de imagens contendo pessoas andando de bicicleta com as bicicletas bem anotadas, além do auxilio do detector de objetos *YOLO* pré-treinado no *COCO* para a detecção automatizada de pessoas, esse algoritmo associa pessoas em posição de condução com sua respectiva bicicleta para gerar anotações de ciclistas, dispensando completamente a necessidade de anotações manuais.

A principal contribuição deste trabalho é um novo conjunto de dados baseado em um sub-conjunto de imagens do *Open Images* contendo anotações de ciclistas. Ao treinar o *YOLO* nesse novo conjunto de dados obtivemos resultados significativamente superiores comparado a quando treinamos esse modelo nos demais conjuntos de dados que contém anotações de ciclistas. Por meio da técnica de transferência de aprendizado [Zhuang et al. 2020], o ambiente escolhido para validar os resultados obtidos foi o Campus Cidade Universitária Armando Salles de Oliveira da Universidade de São Paulo (USP), localizado no bairro do Butantã, em São Paulo, e conhecido como Campus São Paulo - Capital. Esse campus conta com uma larga infraestrutura de monitoramento em tempo real e possui restrições de dias e horários para a prática do ciclismo esportivo. Devido à sua extensão de aproximadamente 3.650.000 m^2, o monitoramento automatizado de ciclistas é uma grande contribuição não somente para a segurança mas também para uma boa convivência de ciclistas, pedestres e automóveis.

O restante deste trabalho está organizado da seguinte forma: na seção 2 apresentamos uma discussão sobre os principais trabalhos relacionados; na seção 3 apresentamos os detalhes do conjunto de dados proposto assim como detalhes de construção; na seção 4 evidenciamos que o conjunto de dados proposto permite o treinamento de detectores de ciclistas com maior poder de generalização para o ambiente da USP, Campus São Paulo - Capital; finalmente, a seção 5 traz as conclusões e possibilidades de trabalhos futuros.

2. Trabalhos Relacionados

Ao longo dos últimos anos foram criados vários conjuntos de dados públicos compostos por milhares de imagens com diversas classes de objetos anotados. Esses conjuntos têm acelerado a evolução de algoritmos de aprendizado profundo para detecção de objetos. Conjuntos de dados como *PASCAL Visual Object Classes (VOC)* [Everingham et al. 2010], *Microsoft COCO* [Lin et al. 2014], *Open Images* [Kuznetsova et al. 2020, Krasin et al. 2017], *MIO-TCD* [Luo et al. 2018], *Tsinghua-Daimler* [Li et al. 2016] e *Specialized Cyclist Detection* [Masalov et al. 2019] contêm imagens com objetos anotados em cenas cotidianas nas quais estes objetos estão presentes no seu contexto natural. Os conjuntos *VOC-PASCAL*, *COCO* e *Open Images* contêm anotação de objetos de classes diversas, enquanto os outros são especializados, com anotações de objetos de uma classe ou um grupo específico de classes. O conjunto *MIO-TCD* é especializado em veículos, para aplicação em análise de tráfego por câmeras de trânsito, e os conjuntos *Tsinghua-Daimler* e *Specialized Cyclist Detection*, em ciclistas, para aplicação em cenários de condução assistida ou autônoma com utilização de câmeras em veículos.

O *VOC-PASCAL* tem 11.540 imagens com 27.450 objetos rotulados em 20 classes, sendo 603 imagens com bicicleta, o *COCO* tem 328 mil imagens com 2,5 milhões de objetos rotulados em 91 classes, sendo cerca de 7 mil imagens com bicicleta e o *Open Images* tem cerca de 9 milhões de imagens com 16 milhões de objetos anotados em 600 classes, sendo cerca de 18 mil imagens com bicicleta. Estes conjuntos de dados costumam ser utilizados em competições para promover o desenvolvimento da visão computacional com métodos de aprendizado profundo, incluindo a detecção de objetos.

O *MIO-TCD* é um conjunto de dados para classificação e localização de veículos, subdividido em imagens para classificação e imagens para detecção de objetos, obtidas

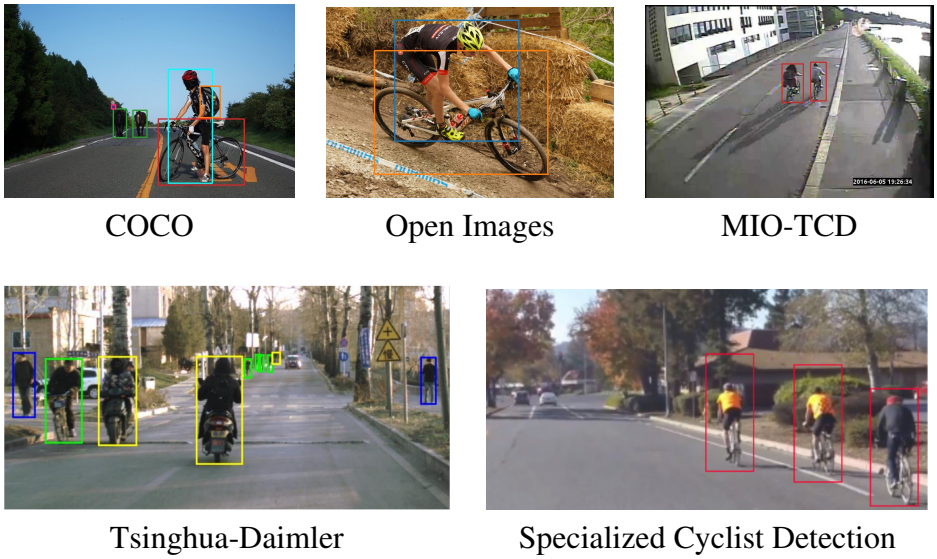

Figura 1. Exemplos de imagens de conjuntos de dados públicos para detecção de objetos, com anotações.

a partir de câmeras de trânsito instaladas no Canadá e nos Estados Unidos. No *MIO-TCD-Classification* as imagens são recortes de cenas contendo um único objeto de interesse e no *MIO-TCD-Localization* as imagens são cenas completas, semelhantes às dos outros conjuntos. Ele contém 137.743 imagens com 416.277 objetos anotados em 12 classes, sendo cerca de 1.933 imagens com bicicleta, cuja anotação inclui também a pessoa que a conduz, coincidindo com a nossa definição de ciclista. Com esse conjunto de dados, Luo et al. treinaram e avaliaram a localização de veículos, nos métodos *Faster R-CNN* [Ren et al. 2015], *SSD* [Liu et al. 2016], *YOLO* [Redmon et al. 2016, Redmon and Farhadi 2017], método de Wang et al. [Wang et al. 2017] e método de Jung et al. [Jung et al. 2017].

O *Tsinghua-Daimler* tem 30.406 imagens no total, com 22.161 ciclistas anotados. Suas imagens foram gravadas por câmera montada em um veículo em movimento no tráfego urbano de Pequim por cerca de 6 horas distribuídas em 5 dias, numa região com alta concentração de ciclistas e pedestres. Xiaofei et al. avaliaram detectores baseados em *ACF* [Dollár et al. 2014], *DPM* [Felzenszwalb et al. 2010] e *R-CNN* [Girshick et al. 2014] com esse conjunto de dados para detecção de ciclistas. Foram três grupos de testes com o conjunto dividido em fácil, moderado e difícil, dependendo do nível de oclusão e proporção do tamanho dos ciclistas em relação à imagem.

O *Specialized Cyclist Detection* tem 62.297 imagens no total, sendo 30 ciclistas diferentes, e cerca de 18.200 ciclistas anotados. As imagens do conjunto de dados foram gravadas por câmera montada em um veículo em dois locais diferentes, com duas condições de tempo e de iluminação diferentes. Esse conjunto contém imagens com nível de detecção fácil, moderado e difícil, definidos pelo nível de oclusão dos ciclistas, porém, Masalov et al. não avaliaram método de detecção com esse conjunto de dados.

Os conjuntos de dados apresentados anteriormente são significativos e fundamentais para o avanço do aprendizado computacional. Na Figura 1 são apresentados exemplos de imagens dos conjuntos de dados citados. Entretanto, os conjuntos com anotação de ci-

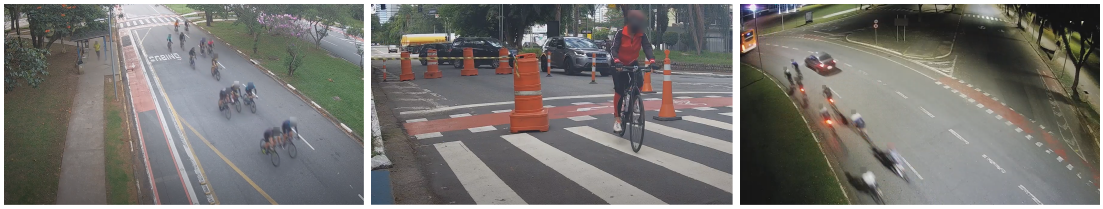

Figura 2. Imagens (devidamente anonimizadas) do Campus São Paulo - Capital (USP).

clistas, contém poucas amostras ou são amostras bastante específicas para seus respectivos domínios de aplicação. Para superar tal limitação, o novo conjunto de dados *Open Images Cyclist*, apresentado neste trabalho, que foi construído a partir do *Open Images*, é mais abrangente para detecção de ciclistas que os conjuntos de dados públicos especializados em ciclistas. As imagens do *Open Images* foram coletadas da comunidade de compartilhamento de imagens *Flickr* e têm diferentes ângulos de observação e variadas escalas e dimensões [Kuznetsova et al. 2020].

3. Conjunto de Dados *OpenImages Cyclists*

O *OpenImages Cyclists* foi criado para viabilizar a detecção de ciclistas em imagens de câmeras de segurança, especialmente no ambiente da USP, Campus São Paulo - Capital. Conforme apresentado na seção 4, esse conjunto de dados não somente permitiu o treinamento de um detector de ciclistas com um desempenho muito bom nas imagens da USP (Figura 2) como também apresentou resultados competitivos nos demais conjuntos de dados contendo ciclistas. Nesta seção, apresentamos os detalhes da construção desse conjunto de dados.

3.1. Construção do Conjunto de Dados

Dentre as quase 18.000 imagens com bicicletas anotadas por caixas delimitadoras disponiveis no *Open Images v6* (em 08/05/2019), cerca de 1/3 apresenta ciclistas esportivos em cenários variados [Krasin et al. 2017]. As imagens desse conjunto contêm muitas bicicletas não anotadas. Além disso, a maioria das anotações das pessoas nessas mesmas imagens são imprecisas ou estão ausentes.

Selecionamos, manualmente, as imagens contendo ciclistas com bicicletas anotadas corretamente. A partir dessa seleção, obtivemos um conjunto com cerca de 6.000 imagens com ciclistas e suas respectivas anotações de bicicleta, desconsiderando nesse momento as anotações de pessoa. Com o auxilio do detector de objetos *YOLOv4*, pré treinado exaustivamente no conjunto de dados *COCO*, automatizamos a criação de novas anotações para as pessoas desse conjunto obtido. Em seguida, fizemos uma avaliação visual por amostragem (cerca de 1200 imagens sorteadas aleatoriamente) para garantir principalmente que as pessoas conduzindo bicicletas foram detectadas satisfatoriamente. De acordo com [Robert et al. 2022], o *YOLOv4* está entre os detectores de melhor desempenho em termos de precisão média sobre o conjunto de dados *COCO*, o qual contém um grande número de instâncias de pessoas anotadas (262.465), mostrando-se suficiente para o nosso experimento.

Para gerar as anotações de ciclistas, uma vez obtidas as anotações de bicicletas e pessoas, é necessário associar corretamente cada condutor com a sua respectiva bici-

(a) Anotação de bicicleta (b) Anotações de pessoas (c) Intersecção entre anotações (d) Anotação de ciclista

Figura 3. Composição das anotações de bicicleta e de pessoa para criar anotação de ciclista (Fonte Open Images). Pessoa ao fundo sem bicicleta (IOU = 0) é descartada no resultado final.

cleta. O algoritmo 1 faz essa composição de anotações automaticamente com base na intersecção entre pessoas e bicicletas, utilizando para isso a métrica da Intersecção sobre a União (IOU, do inglês *Intersection over Union*). Essa técnica consiste em calcular a área de intersecção dividida pela área da união entre duas caixas delimitadoras (pessoa e bicicleta). O resultado é um valor entre zero e um, para o qual zero representa a ausência de intersecção e um representa que uma caixa delimitadora está totalmente inclusa na outra.

O Algoritmo 1, descrito abaixo, inicia recebendo como entrada uma lista de imagens juntamente com as anotações de bicicletas das imagens pré-selecionadas e retorna as anotações de ciclistas para essas mesmas imagens (linha 1). Para cada imagem de entrada, as anotações de bicicletas são colocadas na lista *bikes_bb* (linha 5) e a imagem é submetida à detecção de pessoas (linha 6). Em seguida, para cada bicicleta, o algoritmo percorre a lista de pessoas encontradas nessa imagem e calcula o IOU para todos os pares pessoa e bicicleta (linhas 7 a 16). O maior IOU encontrado muito provavelmente corresponde à pessoa que está conduzindo a bicicleta. Tanto pessoas quanto bicicletas que não tenham intersecção com suas classes complementares, são descartadas (e.g. pessoas ao fundo observando ciclistas ou uma eventual bicicleta sem um condutor). As novas anotações de ciclistas correspondem às mínimas caixas delimitadoras resultantes da união das anotações de cada bicicleta com seu respectivo condutor.

A Figura 3 ilustra o funcionamento desse processo onde a subfigura 3a contém as anotações de bicicleta provenientes do *Open Images*. Já a subfigura 3b é o resultado da detecção de pessoas, contendo a pessoa que conduz a bicicleta e também um expectador. A região sombreada na subfigura 3c corresponde à área de intersecção entre os dois objetos. Finalmente, na subfigura 3d uma nova anotação de ciclista é gerada e a pessoa detectada ao fundo é descartada.

3.2. Visão Geral do Conjunto de Dados

A geração automatizada das anotações de ciclistas resultou em 5.463 imagens com 15.597 instâncias de ciclistas em cenas cotidianas. Cada imagem contém em média três ciclistas, a maioria dos quais tem algum grau de oclusão ou está truncada. Apenas 545 ciclistas aparecem totalmente visíveis. As resoluções das imagens são variadas, por exemplo, 1024

Algoritmo 1 Algoritmo para a geração automatizada de anotações de ciclistas

Entrada: Lista de imagens com bicicletas anotadas
Saída: Lista de imagens com **ciclistas** anotados

```
L ←Lista de imagens com bicicletas anotadas
C ← ∅                                   ▷ Lista de anotações de ciclistas por imagem
for i ∈ L do
    cyc_bb ← ∅
    bikes_bb ← bike_annotations(i)
    persons_bb ← detect_person(i)
    for b ∈ bikes_bb do
        max_iou ← 0
        p_tmp ← ∅
        for p ∈ persons_bb do
            iou ← calculate_iou(b, p)
            if iou > max_iou then
                max_iou ← iou
                p_tmp ← p
            end if
        end for
        if max_iou > 0 and p_tmp ≠ ∅ then
            c ← join_bb(b, p_tmp)
            cyc_bb.append(c)
        end if
    end for
    C.append(i, cyc_bb)
end for
return C
```

x 494, 852 x 768, 1024 x 1024. O tamanho relativo dos ciclistas também é bastante variado, ocupando desde 0,3% até mais de 90% da área da imagem. Durante o treinamento do detector de ciclistas, conforme apresentado na seção 4, esses dados são divididos dinamicamente entre treinamento e teste com uma proporção que pode variar de 80/20 até 90/10 com base na estratégia de *K-Folding*. As anotações de ciclistas produzidas neste trabalho, assim como as imagens utilizadas, estão disponíveis em https://data.ime.usp.br/oic.

4. Resultados Experimentais

Realizamos quatro experimentos para comparar o novo conjunto de dados com outros conjuntos contendo anotações de ciclistas. Apenas as imagens com anotações de ciclistas fizeram parte dos experimentos. Utilizamos 5.463 imagens do *OpenImages Cyclists*, 1.933 imagens do *MIO-TCD-Localization*, 13.655 imagens do *Tsinghua-Daimler* e 7.687 imagens do *Specialized Cyclist Detection*.

O modelo de detecção de objetos escolhido para avaliar o *OpenImages Cyclists* foi o *YOLOv4*. De acordo com [Bochkovskiy et al. 2020], a versão 4 do *YOLO* é apresentada como modelo eficiente de detecção de objetos, criado a partir da composição de métodos de última geração. Comparam com outros detectores de objetos de última geração utilizando o conjunto de dados *COCO*, e consideram que o *YOLOv4* é superior aos detectores

mais rápidos e precisos em termos de taxa de quadros processados (FPS) e precisão média (AP). Ainda segundo [Bochkovskiy et al. 2020], o *YOLOv4* é duas vezes mais rápido que o *EfficientDet* [Tan et al. 2020] com desempenho comparável e melhora o AP e o FPS do *YOLOv3* [Redmon and Farhadi 2018] em 10% e 12%, respectivamente.

Zaidi et al. fizeram uma análise aprofundada dos principais detectores de objetos baseados em aprendizado profundo em [Zaidi et al. 2022], fornecendo uma revisão abrangente deste tipo de detector. Consideraram o *YOLOv4* o estado da arte para detectores de estágio único em tempo real. Avaliaram o desempenho dos modelos com base nos resultados de seus artigos, comparando precisão média e quadros por segundo processados em tempo de inferência.

Considerando o melhor compromisso entre precisão e velocidade dos detectores de objetos avaliados tanto por [Bochkovskiy et al. 2020] quanto por [Zaidi et al. 2022], em nosso experimento utilizamos apenas o *YOLOv4* para avaliar e comparar os conjuntos de dados. Um dos objetivos deste trabalho é fornecer uma solução que possa ser executada em tempo real no ambiente de monitoramento por câmeras da USP.

Com o intuito de avaliar a capacidade de generalização de um detector de ciclistas treinado no novo conjunto, criamos o conjunto de dados *Ciclistas USP* com imagens obtidas a partir das câmeras de segurança da infraestrutura de monitoramento do Campus São Paulo - Capital (USP). Nesse conjunto, 284 imagens com média de 4 ciclistas por imagem foram anotadas manualmente. Suas imagens, semelhantes às ilustradas na Figura 2, em geral, retratam cenas com ciclistas esportivos.

A princípio, a cardinalidade *Ciclistas USP* pode parecer insuficiente para avaliar a capacidade de generalização dos detectores avaliados na seção 4.2. No entanto, essas imagens são uma amostra com uma distribuição uniforme dos dados coletados de junho de 2021 a fevereiro de 2022 por nove câmeras posicionadas em locais e ângulos distintos no campus da USP. Em um trabalho futuro, o conjunto de dados proposto neste trabalho será utilizado para auxiliar a anotação dos demais ciclistas registrados nesse período.

4.1. Metodologia

O primeiro experimento consistiu no treinamento de modelos de detector com a rede padrão do *YOLOv4* em cada um dos conjuntos de dados e posterior avaliação da precisão média (AP), considerando $IOU \geq 0,50$, de cada modelo para detecção de ciclistas em todos os conjuntos de dados. O treinamento foi feito utilizando a técnica de *k-folding* para $k = 5$ (80% dos dados para treinamento e 20% para teste) em 21.000 lotes de 64 imagens, totalizando 266 épocas. Apresentamos os resultados na Tabela 1.

O segundo experimento consistiu na avaliação de AP, considerando $IOU \geq 0,50$, dos detectores para detecção de ciclistas no conjunto de dados *Ciclistas USP*. Esses dados foram utilizados exclusivamente para avaliação da generalização dos detectores, portanto, não foram utilizados para treinamento de nenhum modelo de detecção. Apresentamos os resultados na Tabela 2.

Já no terceiro experimento, treinamos um modelo com todas as imagens de ciclistas dos conjuntos *MIO-TCD-Localization*, *Tsinghua-Daimler* e *Specialized Cyclist Detection* combinadas (23.275 imagens) e avaliamos a detecção de ciclistas no conjunto *Ciclistas USP*, obtendo AP, com $IOU \geq 0,5$, de 65,21%, que é menor do que AP de

detecção de ciclistas, nesse mesmo conjunto, do modelo treinado no *OpenImages Cyclists* (AP = 77.98%)

Complementarmente, no quarto experimento, também utilizando o conjunto de dados *Ciclistas USP* para avaliação, a mudança de $IOU \geq 0,5$ para $IOU \geq 0,75$ causou redução de AP diferente em cada modelo, dependendo do conjunto de dados utilizado para treinamento. Com o *OpenImages Cyclists*, a redução foi de 30%, com o *MIO-TCD-Localization*, de 50%, com o *Tsinghua-Daimler*, de 44% e com o *Specialized Cyclist Detection* de 48%. Esse experimento mostra que o conjunto de dados proposto, além de generalizar melhor do que os demais conjuntos de dados para a detecção de ciclistas, como apresentado na seção 4.2, é também mais preciso.

4.2. Análise de resultados

Para identificação nas Tabelas 1 e 2, nomeamos os detectores treinados com imagens originadas dos conjuntos de dados *Open Images Cyclist*, *MIO-TCD-Localization*, *Tsinghua-Daimler* e *Specialized Cyclist Detection*, respectivamente, como YOLO_{OIC}, YOLO_{MIO}, $\text{YOLO}_{\text{Daimler}}$ e $\text{YOLO}_{\text{Specialized}}$. Para os mesmos conjuntos de dados, identificamos seus respectivos conjuntos de teste por *OIC*, *MIO*, *Daimler* e *Specialized*.

A Tabela 1 apresenta a comparação entre o desempenho dos detectores. Nas linhas estão os detectores e nas colunas, os conjuntos de dados de teste. Cada célula mostra AP do detector indicado na respectiva linha para detecção de ciclistas nas imagens do conjunto de teste indicado na respectiva coluna. A diagonal principal representa os casos nos quais o conjunto de teste e o de treinamento são originários do mesmo conjunto de dados e contém, em geral, os maiores valores de AP, por coluna. Isso é esperado, pois as imagens de ambos os conjuntos têm a mesma distribuição de probabilidade. A exceção do *MIO* ocorre, possivelmente, pela quantidade de ciclistas ser relativamente pequena nos dados de treinamento do YOLO_{MIO}.

Observamos também pela Tabela 1, que o detector YOLO_{OIC} tem melhor AP para detecção de ciclistas em todos os conjuntos de teste, excetuando os casos da diagonal. Esse resultado corrobora a hipótese inicial de que, com a utilização de imagens de boa qualidade, com boa variabilidade e bem anotadas, é possível treinar detectores de objetos mais precisos e, no caso das imagens da USP, resultou em um modelo mais generalizável.

A Tabela 2 apresenta a capacidade de generalização dos detectores para o domínio da USP. Nas colunas estão os detectores e cada célula mostra AP do detector indicado na respectiva coluna para detecção de ciclistas nas imagens do conjunto de dados *Ciclistas USP*. Maior AP do detector YOLO_{OIC} para detecção de ciclistas no conjunto *Ciclistas USP*, reforça a melhor capacidade de generalização do detector YOLO_{OIC}, já indicada pela Tabela 1. Além disso, o bom desempenho do detector YOLO_{OIC} no *Ciclistas USP* permite sua aplicação no monitoramento das vias do Campus São Paulo - Capital da USP para detecção dos ciclistas profissionais.

A menor capacidade de generalização dos outros detectores, possivelmente, pode ser justificada pela pequena quantidade de ciclistas nos dados de treinamento do YOLO_{MIO}, pela pouca variabilidade das cenas nos dados de treinamento do $\text{YOLO}_{\text{Daimler}}$ e do $\text{YOLO}_{\text{Specialized}}$, sendo um pouco maior neste último, e pela diferença de ângulos de observação entre as imagens do *MIO* e *Ciclistas USP* em relação ao *Daimler* e *Specialized*.

Tabela 1. Comparação entre desempenho dos detectores (AP)

	OIC	MIO	Daimler	Specialized
$YOLO_{OIC}$	**86.37%**	**99.72%**	63.19%	82.33%
$YOLO_{MIO}$	35.34%	92.84%	10.85%	59.39%
$YOLO_{Daimler}$	63.02%	27.41%	**73.59%**	81.08%
$YOLO_{Specialized}$	76.93%	36.35%	48.49%	**96.01%**

Tabela 2. Capacidade de generalização dos detectores (AP)

	$YOLO_{OIC}$	$YOLO_{MIO}$	$YOLO_{Daimler}$	$YOLO_{Specialized}$
Conjunto de dados USP	**77.98%**	52.82%	29.36%	40.06%

O novo conjunto de dados *OpenImages Cyclists*, graças ao seu processo de construção, apresenta maior diversidade nas imagens que os outros conjuntos de dados, com variação no ângulo de observação, dimensão do ciclista em relação à imagem, posição, cor, fundo, número de ciclistas. Além disso, suas imagens são de boa qualidade porque, em geral, o *Open Images* tem imagens com grande diversidade de cenas e de qualidade superior aos demais conjuntos de dados públicos em termos de resolução, nitidez e iluminação, pois a comunidade *Flickr* permite essa qualidade de imagem [MacAskill 2018]. Desse modo, espera-se que modelos treinados nesse conjunto de dados resultem em detectores mais acurados e com maior capacidade de generalização do que treinados nos outros conjuntos de dados.

5. Conclusão e Trabalhos Futuros

O novo conjunto de dados *OpenImages Cyclists* aprimorou substancialmente a precisão da detecção dos ciclistas nas imagens das câmeras de segurança da USP, resultado potencialmente replicável em ambientes de monitoramento análogos. A variabilidade dos dados de treinamento, expressa pela variação de posicionamento da câmera, iluminação, e a qualidade das imagens favorecem a generalização da detecção.

Um detector de objetos baseado em Aprendizado Profundo e treinado nesse conjunto de dados certamente contribui para aumentar a segurança de ciclistas, os quais muitas vezes necessitam compartilhar as vias com automóveis e pedestres. O bom desempenho do detector *YOLOv4*, treinado no *OpenImages Cyclists*, quando avaliado nos conjuntos de dados *Tsinghua-Daimler* e *Specialized Cyclist Detection* é um forte indício de que nosso conjunto de dados também poderia ser aplicado no contexto de condução autônoma.

Os próximos passos deste trabalho incluem: 1) a detecção de pelotão de ciclistas, para a qual o conjunto de Dados *OpenImages Cyclists* será de fundamental importância; 2) a criação de nova anotação para classes de objetos compostos por outros objetos, como por exemplo motociclistas, utilizando o mesmo processo para criação da anotação de ciclistas.

Agradecimentos

Agradecemos as agências de fomento que financiam nossas atividades de pesquisa: FAPESP nro. 2020/06950-4 (Center for Research and Development on Live Knowledge) e CNPq 308820/2021-5

Referências

[Bochkovskiy et al. 2020] Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *ArXiv*, abs/2004.10934.

[Dollár et al. 2014] Dollár, P., Appel, R., Belongie, S., and Perona, P. (2014). Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1532–1545.

[Everingham et al. 2010] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338.

[Felzenszwalb et al. 2010] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645.

[Girshick et al. 2014] Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 580–587. IEEE.

[Jung et al. 2017] Jung, H., Choi, M.-K., Jung, J., Lee, J.-H., Kwon, S., and Jung, W. Y. (2017). Resnet-based vehicle classification and localization in traffic surveillance systems. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 934–940. IEEE.

[Krasin et al. 2017] Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Veit, A., et al. (2017). Openimages: A public dataset for large-scale multi-label and multi-class image classification. https://github.com/openimages.

[Kuznetsova et al. 2020] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., et al. (2020). The open images dataset v4. *International Journal of Computer Vision*, 128(7):1956–1981.

[Li et al. 2016] Li, X., Flohr, F., Yang, Y., Xiong, H., Braun, M., Pan, S., Li, K., and Gavrila, D. M. (2016). A new benchmark for vision-based cyclist detection. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, pages 1028–1033. IEEE.

[Lin et al. 2014] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

[Liu et al. 2016] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European Conference on Computer Vision*, pages 21–37. Springer.

[Luo et al. 2018] Luo, Z., Branchaud-Charron, F., Lemaire, C., Konrad, J., Li, S., Mishra, A., Achkar, A., Eichel, J., and Jodoin, P.-M. (2018). Mio-tcd: A new benchmark dataset for vehicle classification and localization. *IEEE Transactions on Image Processing*, 27(10):5129–5141.

[MacAskill 2018] MacAskill, D. (2018). Putting your best photo forward: Flickr updates. https://blog.flickr.net/.

[Masalov et al. 2019] Masalov, A., Matrenin, P., Ota, J., Wirth, F., Stiller, C., Corbet, H., and Lee, E. (2019). Specialized cyclist detection dataset: Challenging real-world computer vision dataset for cyclist detection using a monocular rgb camera. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 114–118. IEEE.

[Redmon et al. 2016] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788. IEEE.

[Redmon and Farhadi 2017] Redmon, J. and Farhadi, A. (2017). Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6517–6525. IEEE.

[Redmon and Farhadi 2018] Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *ArXiv*, abs/1804.02767.

[Ren et al. 2015] Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, volume 28, pages 91–99. Curran Associates, Inc.

[Robert et al. 2022] Robert, Ross, Marcin, Elvis, Guillem, Andrew, and Thomas (2022). Papers with code. https://paperswithcode.com/sota/object-detection-on-coco. Acessado em 20/05/2022.

[Santhosh et al. 2020] Santhosh, K. K., Dogra, D. P., and Roy, P. P. (2020). Anomaly detection in road traffic using visual surveillance: A survey. *ACM Comput. Surv.*, 53(6).

[Tan et al. 2020] Tan, M., Pang, R., and Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10778–10787. IEEE.

[Wang et al. 2017] Wang, T., He, X., Su, S., and Guan, Y. (2017). Efficient scene layout aware object detection for traffic surveillance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 926–933. IEEE.

[Zaidi et al. 2022] Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., and Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126:103514.

[Zhang et al. 2021] Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Commun. ACM*, 64(3):107–115.

[Zhou et al. 2017] Zhou, X., Gong, W., Fu, W., and Du, F. (2017). Application of deep learning in object detection. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pages 631–634. IEEE.

[Zhuang et al. 2020] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76.