Um Estudo de Caso da Detecção do Uso de Máscaras Faciais com Redes Neurais Convolucionais Regionais

Diego Lucena de Medeiros, Elloá B. Guedes, Carlos Maurício S. Figueiredo

[1]Grupo de Pesquisa em Sistemas Inteligentes
Universidade do Estado do Amazonas (UEA)
Av. Darcy Vargas, 1200 – Manaus – Amazonas
{dldm.eng18,ebgcosta,cfigueiredo}@uea.edu.br

Abstract. *Aiming at supporting the development of solutions for Smart Cities in preventing the spread of airborne spread infectious diseases whose prevention strategies rely on the use of facial masks, this work considered the Computer Vision face mask detection problem with YOLOv3 and YOLOv5 models in a case study with three different realistic datasets. Experimental results highlighted the YOLOv5 Small 6 as the reference solution with a mAP of* 92.8% *in a validation scenario with unified examples. We also performed transfer learning on such model with images from AIZOO dataset and compared the performance with solutions from literature. We verified that the proposed model is competitive with state-of-art alternatives and has a strong potential to be embedded in low-resources computational devices.*

Resumo. *Com o objetivo de desenvolver soluções para Cidades Inteligentes que colaborem na mitigação da propagação de doenças, este trabalho considerou o problema de Visão Computacional de detecção do uso de máscaras faciais, o qual foi abordado com os modelos YOLOv3 e YOLOv5 em um estudo de caso com três conjuntos de dados distintos e realísticos. Os resultados experimentais destacaram o YOLOv5 Small 6 como a solução de referência com um mAP de* $92{,}8\,\%$ *em um cenário de validação com exemplos unificados. Também foi realizada uma transferência de aprendizado desse modelo com imagens do conjunto de dados AIZOO e os resultados foram comparados com soluções da literatura, em que verificou-se competitivo com as alternativas do estado da arte e com forte potencial para ser embarcado em dispositivos computacionais com recursos limitados.*

1. Introdução

De acordo com a Organização Mundial de Saúde (OMS), a pandemia do COVID-19, causada pelo vírus SARS-CoV-2, emergiu como um fardo avassalador para a saúde em todo o planeta, com característica altamente infecciosa com transmissão via aérea, por meio da propagação de aerossóis e gotículas [OMS 2019]. Até a ocasião da conclusão do trabalho, em meados de maio de 2022, registros denotavam $524.777.204$ casos e $6.281.495$ óbitos em âmbito global [JHU 2019]. A contaminação do SARS-CoV-2 pode ser reduzida por meio do distanciamento social e por protocolos de higienização [Srivastava et al. 2020]. O uso de máscaras faciais, em particular, é também uma estratégia para mitigar o contágio [Prather et al. 2020], a qual mostrou-se comprovadamente efetiva ainda que o distanciamento social não pudesse ser assegurado [Kwon et al. 2021].

Segundo a Organização das Nações Unidas (ONU), mais da metade da população mundial vive em áreas urbanas ($55\,\%$) e projeta-se que no ano de 2050 esta proporção seja de

68 % [ONU 2018]. Além dos problemas esperados decorrentes desse processo de urbanização, novas doenças altamente contagiosas podem surgir em áreas densamente povoadas. Na verdade, o potencial negativo desse padrão urbano foi evidenciado na pandemia de COVID-19, que se espalhou rapidamente em todos os continentes em apenas algumas semanas. Para evitar ou mesmo reduzir os impactos desta e das próximas pandemias, a gestão da informação pode ser tão vital quanto qualquer outro elemento dos sistemas de Saúde, e os dados fornecidos pelas cidades são cruciais nesse aspecto [Costa and Peixoto 2020].

Uma Cidade Inteligente (CI) é uma área urbana que usa a informação coletada por diferentes tipos de sensores para monitorar e administrar os recursos disponíveis de forma eficiente, com capacidade de aprender e se adaptar continuamente, melhorando as condições de conforto e segurança para as pessoas que nela habitam [Du et al. 2019]. No entanto, como parte da preparação eficaz para as pandemias atuais e futuras, espera-se que o desenvolvimento sustentável das CIs forneça inteligência situacional e uma resposta direcionada automatizada para garantir a segurança da saúde pública [Shorfuzzaman et al. 2021].

Visando colaborar na perspectiva do desenvolvimento de soluções que auxiliem as CIs a mitigar a disseminação de doenças infecciosas transmitidas pelo ar (influenza, MERS, COVID-19, etc.) cujas estratégias de prevenção contemplem o uso de máscaras faciais, este trabalho apresenta um estudo de caso do uso de Redes Neurais Convolucionais Regionais Profundas (R-CNNs, do inglês *Regional Convolutional Neural Networks*) da família YOLO (do inglês, *You Only Look Once*) para localização e classificação de sujeitos com e sem máscaras faciais em diferentes contextos. Para tanto, foram consideradas bases de dados realísticas da literatura, as quais continham exemplos de sujeitos de diferentes faixas etárias, em ambientes não controlados e fazendo uso de diferentes tipos de máscaras. Os resultados obtidos, aferidos em contextos experimentais sob diferentes métricas de desempenho, apontam o potencial dos modelos investigados em tais aplicações práticas para CIs.

Para apresentar os resultados obtidos, este trabalho está organizado como segue. Uma visão geral dos trabalhos relacionados encontra-se na Seção 2. As bases de dados utilizadas, modelos e suas configurações bem como as estratégias de avaliação das soluções propostas encontram-se descritos na Seção 3. Os resultados obtidos são apresentados, contrastados e discutidos na Seção 4. Por fim, as considerações finais e sugestões de trabalhos futuros são apresentadas na Seção 5.

2. Trabalhos Relacionados

Soluções automáticas para o problema do monitoramento do uso de máscaras faciais em populações consistem em algoritmos de Visão Computacional que abordam a localização de uma ou mais faces em uma imagem, ou em uma sequência de *frames* de vídeo por meio da estimação de caixas delimitadoras (*bouding boxes*). Em seguida, produz-se uma classificação binária para cada caixa quanto ao uso de máscara na respectiva face, por exemplo, "com máscara" (classe positiva) ou "sem máscara" (classe negativa). A combinação da localização e da classificação compõem a tarefa de detecção do uso de máscaras faciais.

De maneira geral, as soluções mais proeminentes na literatura para a detecção do uso de máscaras faciais baseiam-se no uso de Redes Neurais Convolucionais Profundas (CNNs, do inglês *Convolutional Neural Networks*) ou em abordagens híbridas, inevitavelmente com alguma abordagem de *Deep Learning* como parte da solução [Nowrin et al. 2021]. Há apenas um único registro de solução na literatura inteiramente baseada na extração de características

com algoritmos de Visão Computacional combinado ao uso de algoritmos tradicionais de *Machine Learning* [Nieto-Rodríguez et al. 2015]. O uso de *Deep Learning* nesse contexto é uma consequência positiva dos benefícios observados dos métodos e técnicas dessa sub-área da Inteligência Artificial no desempenho de diversas tarefas de Visão Computacional, ao promover o aprendizado massivo de características hierárquicas sucessivamente complexas [Goodfellow et al. 2016, Khan et al. 2018].

No escopo da detecção de objetos com *Deep Learning*, tem-se a proposição de regiões de interesse e a obtenção das respectivas classificações com uma CNN, compondo as R-CNNs. Quando estas duas etapas são feitas em um único estágio, têm-se os detectores *single-shot*, como é o caso dos algoritmos da família YOLO [Redmon et al. 2016]. Nesta última, uma imagem de entrada é dividida em múltiplas regiões e há a previsão simultânea das caixas delimitadoras e das probabilidades de classificação de cada região, o que favorece a sua utilização para detecção de objetos em tempo real.

No tocante à utilização da YOLO para detecção do uso de máscaras faciais, Loey *et al.* [2021] consideraram uma combinação da CNN ResNet-50, para extração de características, com a YOLOv2, uma melhoria subsequente da YOLO originalmente proposta [Redmon and Farhadi 2017]. Ao abordarem a detecção de máscaras médicas ou cirúrgicas, os autores reportaram uma precisão de $81\,\%$ em um cenário de avaliação experimental.

A YOLOv3 decorreu de melhorias no projeto da YOLOv2, tal como a utilização da DarkNet-53, inspirada na CNN ResNet, a qual contém 53 camadas convolucionais e também camadas residuais, combinação considerada mais eficaz e eficiente para extração de características. Prevê um tensor tridimensional que codifica as caixas delimitadoras, as probabilidades de nelas haver objetos e também as previsões das respectivas classes [Redmon and Farhadi 2018]. O trabalho de Singh *et al.* fez uso da YOLOv3 em comparação com modelos de detecção em dois estágios em um cenário experimental com duas bases de dados abertas complementadas por exemplos coletados e rotulados pelos próprios autores. Os resultados obtidos destacaram a YOLOv3 com AP (*Average Precision*) de $55\,\%$ com tempo de inferência igual a $0{,}045\,\mathrm{s}$, o que ressalta o potencial para detecção em tempo real [Singh et al. 2021].

Há trabalhos mais recentes na literatura que consideram a YOLOv4 [Mahurkar and Gadge 2021] e a YOLOv5 [Yang et al. 2020] no problema em questão, também evidenciando a qualidade de tais modelos na detecção de máscaras. A YOLOv4 incorporou uma nova estratégia de aumento artificial de dados, técnicas de regularização e treinamento com aceleração em hardware via GPU [Bochkovskiy et al. 2020]. Já a YOLOv5, por sua vez, foi projetada para dar prosseguimento às melhorias na YOLOv4, facilitando especialmente os aspectos de implementação tecnológica, reduzindo o tempo de treinamento bem como o número de parâmetros, ainda promovendo uma alta acurácia [Jocher et al. 2020].

Apesar de tais registros e progressos, conforme argumentam Nowrin *et al.* em um *survey* de bases de dados e soluções para o problema da detecção de máscaras faciais [Nowrin et al. 2021], há uma grande variedade nas bases de dados públicas e privadas na avaliação experimental das soluções disponíveis na literatura, o que dificulta uma comparação mais objetiva dentre as proposições e que ressalta a importância de um *benchmark dataset* com variedade de exemplos compatíveis com a detecção em contextos reais. Neste sentido, visando explorar uma análise comparativa entre diferentes modelos da família YOLO sob diferentes bases de dados publicamente disponíveis na literatura para o problema de detecção de máscaras faciais

é que o estudo de caso deste trabalho foi delineado, cuja metodologia é descrita a seguir.

3. Materiais e Métodos

O problema considerado no escopo deste trabalho foi abordado como uma tarefa de detecção mediante Aprendizado Supervisionado. Os dados experimentais, modelos e a avaliação de desempenho são descritos detalhadamente nas subseções a seguir.

3.1. Dados Experimentais

Os dados experimentais utilizados no escopo deste trabalho são oriundos de bases de dados publicamente disponíveis relacionadas ao problema de detecção do uso de máscaras. Primeiramente, foram pré-selecionadas 60 bases de dados relacionadas ao problema em repositórios como Kaggle, IEEEDataport e Mendeley Data, as quais foram analisadas quanto aos critérios do tipo de anotação (rótulos para detecção), de classes (binária, com e sem máscara), qualidade e quantidade das imagens. Dessa análise, três bases de dados foram então selecionadas, as quais são apresentadas a seguir:

1. **Mask Dataset (Mask)**. Base de dados com 848 imagens rotuladas para detecção contendo exemplos de indivíduos com ou sem máscaras faciais. Além das duas classes desejadas, possuía exemplos de utilização incorreta de máscara (21 imagens), os quais foram excluídos para os propósitos deste trabalho. Os exemplos disponíveis são realísticos e contemplam sujeitos de diversas idades, em diversas posições e com tipos diferentes de máscara (de tecido, PFF2, etc.). Os autores da base de dados não sugerem uma partição dos exemplos em treino, teste e validação [MakeML 2020];
2. **Face Mask Dataset (Face Mask)**. Base de dados de detecção com 920 exemplos rotulados e particionados em conjuntos de treino (700 imagens), validação (100 imagens) e teste (120 imagens). Segundo o autor, os exemplos foram coletados do resultado da consulta de palavras-chaves em mecanismos de busca na internet, além da junção com exemplos de outras bases de dados livremente disponíveis. Cada imagem dessa base de dados costumam conter mais de um sujeito [Purohit 2020];
3. **MaskDetection at YOLO format (Mask Detection)**. Base de dados com 1.226 imagens previamente repartidas em conjuntos de treino, validação e teste com 768, 298 e 160 exemplos, respectivamente. As imagens foram coletadas com o uso de ferramentas de *web scraping* em mecanismos de busca. A maior parte das imagens possui apenas uma anotação, sendo ela de um sujeito com ou sem máscara [Lorenzo 2020];

As bases de dados escolhidas são realísticas e representativas, pois apresentam exemplos nas mais diversas condições de iluminação, resolução, etc., conforme ilustrado na Figura 1. Uma análise descritiva dos rótulos de detecção das três bases de dados é apresentada na Tabela 1.

Além do uso individual de tais bases, foi elaborada uma *Base Unificada*, a qual contemplou a união de todos os exemplos disponíveis nas três bases de dados previamente mencionadas. Após a união dos 2.994 exemplos, os mesmos foram então sujeitos a uma escolha randomizada e distribuição em partições de treino, validação e teste contendo $70\,\%$, $10\,\%$ e $20\,\%$ dos exemplos, respectivamente.

(a) Mask (b) Face Mask (c) Mask Detection

(d) Mask (e) Face Mask (f) Mask Detection

Figura 1: Exemplos de diferentes imagens oriundas das três bases de dados.

Tabela 1: Descrição estatística dos *pixels* das *bounding boxes* nas bases de dados selecionadas.

	Comprimento			Largura		
	Média e Desvio Padrão	**Máx**	**Mín**	**Média e Desvio Padrão**	**Máx**	**Mín**
Mask	34.95 ± 32.64	340	2	31.07 ± 27.91	317	1
Face Mask	95.34 ± 108.93	2016	8	83.09 ± 89.24	1612	7
Mask Detection	163.29 ± 140.42	847	11	165.04 ± 139.48	1337	10

3.2. Arquiteturas, Parâmetros e Hiperparâmetros

As arquiteturas da família YOLO escolhidas para o escopo deste trabalho foram a YOLOv3 [Redmon and Farhadi 2018], por ser uma melhoria continuada proposta pelos mesmos autores da primeira versão dessa solução, e a YOLOv5 [Jocher et al. 2020], por ser o estado da arte dessa família de detectores *single-shot*. Foram escolhidas variantes destas arquiteturas considerando diferentes quantidades de parâmetros, conforme disposto na Tabela 2.

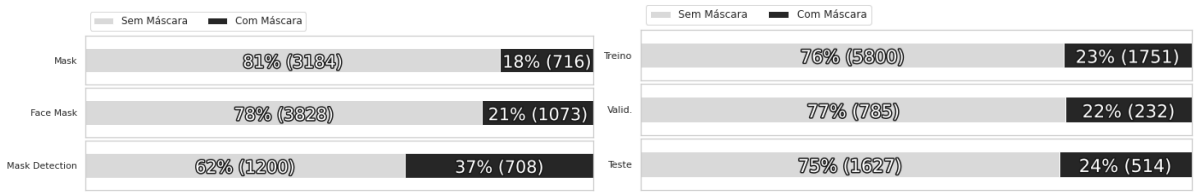
Em todas as arquiteturas os pesos foram inicializados de forma aleatória, sem o uso de estratégias de transferência de aprendizado; foram utilizadas 1.000 épocas máximas para o treinamento; regularização com *early stopping*; persistência em disco do melhor conjunto de pesos durante o treinamento perante o conjunto de validação (*model checkpoint*); *batches* com 32 imagens; e taxa de aprendizado adaptativa. Os demais parâmetros e hiperparâmetros foram mantidos em seus valores padrão.

Tabela 2: Descrição dos parâmetros dos modelos.

Variantes YOLOv3		Variantes YOLOv5	
Modelo	**Parâmetros**	**Modelo**	**Parâmetros**
YoloV3 Small	8.669.002	**YoloV5 Nano 6**	1.761.871
YoloV3	61.502.815	**YoloV5 Small 6**	7.015.519
		YoloV5 Medium 6	20.856.975

3.3. Avaliação de Desempenho

A avaliação de desempenho dos modelos foi efetuada segundo uma validação cruzada do tipo *holdout* para cada base de dados, da seguinte forma: para cada modelo, o mesmo foi treinado com as partições de treino e validação (para fins de regularização), tendo seu desempenho posteriormente aferido a partir das previsões efetuadas para a partição de testes. A quantidade de exemplos por classe nas bases consideradas da literatura e na base unificada são apresentados na Fig 2. Ressalta-se que a base unificada mostra-se proporcional no tocante à quantidade de exemplos por classe nas três partições existentes.



(a) Bases Individuais **(b) Base Unificada**

Figura 2: Propoção de exemplos por classe nas bases de dados.

O desempenho dos modelos foi aferido no tocante à correta classificação e ao grau de sobreposição das *bounding boxes* em comparação com os rótulos disponíveis na partição de testes, o qual foi sintetizado conforme métricas de tarefas de detecção: precisão, revocação, F_1-*Score*, AP (do inglês, *Average Precision*) e mAP (do inglês *Mean Average Precision*), em que nesta última considerou-se o limiar do IoU (do inglês, *Intersection Over Union*) como sendo maior igual a $0{,}5$. Uma explicação detalhada do cálculo 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 das redes YOLO 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 teste para as diferentes variantes das arquiteturas encontram-se dispostos na Tabela 3, acrescidos do número de *Frames* Por Segundo (FPS).

Ao observar detalhadamente os valores das métricas, percebeu-se que as variantes das arquiteturas YOLOv3 e YOLOv5 abordaram a contento o problema da detecção proposto no

Tabela 3: Avaliação de desempenho dos modelos perante as bases de dados consideradas.

	Base de Dados	Precisão	Revocação	F_1-*Score*	AP_0	AP_1	mAP	FPS
YoloV3 Small	**Mask**	88,6	72,9	80,0	72,1	89,2	80,7	37,3
	Face Mask	90,1	81,9	85,8	78,4	94,5	86,4	37,0
	Mask Detection	96,3	96,7	96,5	99,2	98,8	99,0	83,3
	Base Unificada	91,9	84,9	88,3	84,3	94,2	89,3	107,5
YoloV3	**Mask**	95,1	79,6	86,7	84,1	93,2	88,7	32,8
	Face Mask	93,6	87,5	90,4	87,6	97,2	92,4	38,3
	Mask Detection	99,8	99,3	99,5	99,5	99,5	99,5	35,5
	Base Unificada	94,2	90,3	92,2	89,5	95,8	92,7	40,2
YoloV5 Nano 6	**Mask**	93,6	75,2	83,4	77,1	90,5	83,8	38,0
	Face Mask	87,5	83,3	85,3	78,1	95,2	86,7	37,3
	Mask Detection	97,2	97,2	97,2	99,4	98,4	98,9	87,0
	Base Unificada	91,6	87,5	89,5	88,0	94,7	91,3	123.5
YoloV5 Small 6	**Mask**	93,3	77,5	84,7	81,2	93,6	87,4	37,3
	Face Mask	90,4	89,2	89,8	85,3	97,5	91,4	35,8
	Mask Detection	99,2	98,5	98,8	99,5	99,1	99,3	77,5
	Base Unificada	94,0	90,1	92,0	89,8	95,9	**92,8**	101.0
YoloV5 Medium 6	**Mask**	91,2	79,9	85,2	80,5	92,5	86,5	35,6
	Face Mask	90,2	85,2	87,6	83,8	93,3	88,6	32,7
	Mask Detection	99,3	98,5	98,9	99,5	99,4	99,4	56,0
	Base Unificada	92,2	90,5	91,3	89,0	95,5	92,3	78,7

estudo de caso, com mAP mínimo igual a $80{,}7\,\%$, máximo igual a $99{,}5\,\%$ e médio igual a $91{,}3\,\% \pm 5{,}4$. Uma das hipóteses para o bom desempenho das YOLO diz respeito à baixa complexidade do padrão a ser aprendido para detectar faces com e sem máscara, especialmente no segundo caso quando se recapitulam os filtros convolucionais para extrair as características de Haar presentes no tradicional algoritmo de detecção de faces de Viola e Jones [Viola and Jones 2001].

De maneira geral, observou-se que o melhor desempenho de todas as redes foi na base de dados *Mask Detection*, com mAP médio de $99{,}2\,\% \pm 0{,}23$, o que sugere que a mesma pode não ser suficientemente representativa para a real complexidade do problema, por exemplo, não contemplando irregularidades na iluminação; oclusão parcial de face; presença de acessórios (óculos, chapéu, etc.) ou de modificações corporais (tatuagens, *piercings*, etc.); diferentes resoluções das imagens; e até mesmo uma baixa diversidade de sujeitos com diferentes cores de pele conforme escala de Fitzpatrick.

A partir dos resultados obtidos também foi possível observar que os maiores valores de FPS em cada arquitetura ocorreram na Base Unificada, o que sugere que o maior quantitativo de exemplos de treino neste cenário contribuiu para um aprendizado mais eficiente da localização das faces. Levando isto em consideração em virtude do potencial de uso em soluções de tempo real juntamente com o melhor desempenho na detecção, ressalta-se a YOLOv5 *Small* 6 como a solução de referência para este estudo de caso. Esta rede obteve mAP igual a $92{,}8\,\%$ na Base Unificada. Uma característica positiva da solução de referência é o seu baixo número de parâmetros, o que pode favorecer a sua implantação em dispositivos móveis ou embarcados, facilitando seu uso em diversos contextos práticos de monitoramento. A Fi-

gura 3 ilustra algumas detecções feitas por esta solução em exemplos presentes no conjunto de testes. Os rótulos alvo da detecção são denotados na cor azul, as previsões para faces sem máscara encontram-se na cor vermelha e as previsões para faces com máscara são denotadas na cor verde. Percebe-se que as principais limitações da solução proposta residem na detecção de exemplos muito pequenos e na ligeira discrepância entre as *bouding boxes* desejadas e previstas.

Figura 3: Exemplos de detecções com a YOLOv5 *Small* 6 no conjunto de testes.

Considerando que não há *benchmarks* bem estabelecidos para avaliação comparativa de soluções de detecção do uso de máscaras faciais, tomou-se então a solução de referência pré-treinada com a Base Unificada e fez-se uma Transferência de Aprendizado com a continuidade do treinamento perante exemplos da base de dados AIZOO, resultante da mesclagem das bases WIDER FACE [Yang et al. 2016] (50 % dos exemplos, contemplando apenas pessoas sem máscara em diversas situações) e MAsked FAces (MAFA) [Ge et al. 2017] (50 %

dos exemplos, apenas sujeitos com máscara), acrescida de rótulos de detecção pelos seus proponentes [AIZOOTech 2022]. Uma vez que as bases de dados para detecção de máscaras faciais ainda não estão bem estabelecidas e há discussões sobre a versatilidade das mesmas, uma boa prática é combinar diferentes bases para aumentar diversidades, número de exemplos e minimizar vieses [Nowrin et al. 2021], o que foi preconizado na Base Unificada e também é considerado pela AIZOO. Além disso, a base AIZOO é utilizada pela literatura conforme partição *holdout* com $77\,\%$ dos exemplos para treino e $23\,\%$ para teste [Fan et al. 2021], o que viabiliza um comparativo mais objetivo com o estado da arte. A Transferência de Aprendizado na solução de referência considerou esta mesma prática com os dados e os resultados obtidos dos testes encontram-se dispostos Tabela 4, juntamente com o comparativo de outras soluções no mesmo contexto.

Tabela 4: Comparativo de desempenho perante a partição de testes da AIZOO.

Modelo	AP_0	AP_1	mAP
RetinaFace [Deng et al. 2020]	92,8	93,1	93,0
RetinaFaceMask-M [Fan and Jiang 2021]	93,6	90,4	92,0
SL-FMDet [Fan et al. 2021]	93,6	94,0	93,8
Solução Proposta	89,4	97,0	93,2

Observa-se que o mAP da solução proposta é competitivo com soluções do estado da arte sob as mesmas condições de avaliação. Em comparação com a RetinaFace, destaca-se que o mAP obtido foi maior, o que é relevante no tocante ao número de parâmetros, especialmente em contraste com o *backbone* computacionalmente custoso deste trabalho relacionado. A RetinaFaceMask-M, baseada na CNN MobileNet, já demonstrava uma baixa capacidade na detecção das máscaras faciais, vide o baixo valor do AP_1, e foi natural que a solução proposta superasse o desempenho deste trabalho relacionado. Por fim, em comparação com o SL-FMDet, solução do estado da arte baseada no uso de redes residuais de atenção e mapas de calor Gaussianos, observa-se que a solução proposta possui desempenho $0{,}63\,\%$ inferior, mas que é $3{,}19\,\%$ superior no tocante à detecção de faces com máscaras. O grande diferencial da SL-FMDet em comparação à solução de referência reside no número de parâmetros ($0{,}43$ M versus $7{,}01$ M) e de operações ($1{,}01$ GFLOPs versus $15{,}8$ GFLOPs). Por outro lado, a solução de referência pode ser facilmente reproduzida com bases de dados públicas e abertas e com bibliotecas *open-source*.

5. Considerações Finais

Este trabalho apresentou os resultados experimentais da avaliação de cinco arquiteturas de R-CNNs da família YOLO perante o problema de detecção do uso de máscaras faciais. Em uma primeira etapa, considerou-se três bases de dados públicas e gratuitas disponíveis na literatura e também a junção das mesmas em uma base unificada para treinar e avaliar os modelos considerados. Como solução de referência foi possível identificar a YOLOv5 *Small* 6 treinada e testada com a Base Unificada obtendo F_1-*Score* igual a $92\,\%$, mAP igual a $92{,}8\,\%$ e capacidade de detecção em tempo real a 101 FPS. O baixo número de parâmetros desta arquitetura propicia a sua implantação em dispositivos com recursos de hardware limitados, favorecendo a implementação da solução em contextos reais de CIs, auxiliando a mitigar a disseminação

de doenças infecciosas transmitidas pelo ar cujas estratégias de prevenção contemplem o uso de máscaras faciais.

Em uma segunda etapa de avaliação da solução proposta perante a base de dados AIZOO mediante Transferência de Aprendizado para fins de comparação com a literatura, verificou-se que o desempenho do modelo aqui proposto é menos de $1\,\%$ inferior ao estado da arte na detecção de máscaras faciais, mas que é mais eficiente para detectar faces com máscara que esta contrapartida. Dessa forma, apresenta-se como uma solução competitiva para o problema da detecção automática de máscaras faciais.

Em trabalhos futuros almeja-se efetuar ajustes finos na solução de referência com vistas a maximizar o seu desempenho perante a Base Unificada, a AIZOO e outras que vierem a ser disponibilizadas na literatura como *benchmark*. Além disso, deseja-se avaliar a solução no contexto da detecção em tempo real em vídeos de câmeras de monitoramento, verificando se o desempenho se mantém consistente nestes cenários práticos, o que pode ser especialmente útil para a utilização em CIs.

Agradecimentos

Os autores agradecem o apoio financeiro da Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) por meio do programa PAIC 2020-2021 e 2021-2022. Agradecem também o apoio material do Laboratório de Sistemas Inteligentes (LSI) da Universidade do Estado do Amazonas.

Referências

AIZOOTech (2022). Face mask detection. Disponível em `https://github.com/AIZOOTech/FaceMaskDetection`. Acesso em 18 de maio de 2022.

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. Disponível em `https://arxiv.org/abs/2004.10934`. Acesso em 18 de maio de 2022.

Costa, D. G. and Peixoto, J. P. J. (2020). COVID-19 pandemic: a review of smart cities initiatives to face new outbreaks. *IET Smart Cities*, 2(2):64–73.

Deng, J., Guo, J., Ververas, E., Kotsia, I., and Zafeiriou, S. (2020). RetinaFace: Single-Shot Multi-Level face localisation in the wild. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Du, R., Santi, P., Xiao, M., Vasilakos, A. V., and Fischione, C. (2019). The sensable city: A survey on the deployment and management for smart city monitoring. *IEEE Commun. Surv. Tutor.*, 21(2):1533–1560.

Fan, X. and Jiang, M. (2021). Retinafacemask: A single stage face mask detector for assisting control of the covid-19 pandemic. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 832–837.

Fan, X., Jiang, M., and Yan, H. (2021). A deep learning based light-weight face mask detector with residual context attention and gaussian heatmap to fight against COVID-19. *IEEE Access*, 9:96964–96974.

Ge, S., Li, J., Ye, Q., and Luo, Z. (2017). Detecting masked faces in the wild with lle-cnns. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 426–434. Disponível em 10.1109/CVPR.2017.53. Acesso em 18 de maio de 2022.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

JHU (2019). Johns Hopkins University (JHU) COVID-19 dashboard (COVID-19) pandemic. Disponível em https://coronavirus.jhu.edu/map.html. Acesso em 18 de maio de 2022.

Jocher, G., Stoken, A., Borovec, J., NanoCode012, ChristopherSTAN, Changyu, L., Laughing, tkianai, Hogan, A., lorenzomammana, 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 18 de maio de 2022.

Khan, S., Rahmani, H., Shah, S., and Bennamoun, M. (2018). *A Guide to Convolutional Neural Networks for Computer Vision*. Number 1 in Synthesis Lectures on Computer Vision. Morgan & Claypool Publishers.

Kwon, S., Joshi, A. D., Lo, C.-H., Drew, D. A., Nguyen, L. H., Guo, C.-G., Ma, W., Mehta, R. S., Shebl, F. M., Warner, E. T., Astley, C. M., Merino, J., Murray, B., Wolf, J., Ourselin, S., Steves, C. J., Spector, T. D., Hart, J. E., Song, M., VoPham, T., and Chan, A. T. (2021). Association of social distancing and face mask use with risk of COVID-19. *Nat. Commun.*, 12(1):3737.

Loey, M., Manogaran, G., Taha, M. H. N., and Khalifa, N. E. M. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustain. Cities Soc.*, 65(102600):102600.

Lorenzo, A. (2020). MaskDetection at YOLO format. Disponível em https://www.kaggle.com/alexandralorenzo/maskdetection/version/6. Acesso em 18 de maio de 2022.

Mahurkar, R. R. and Gadge, N. G. (2021). Real-time covid-19 face mask detection with YOLOv4. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE.

MakeML (2020). Mask dataset. Disponível em https://makeml.app/datasets/mask. Acesso em 18 de maio de 2022.

Nieto-Rodríguez, A., Mucientes, M., and Brea, V. M. (2015). System for medical mask detection in the operating room through facial attributes. In *Pattern Recognition and Image Analysis*, Lecture notes in computer science, pages 138–145. Springer International Publishing, Cham.

Nowrin, A., Afroz, S., Rahman, M. S., Mahmud, I., and Cho, Y.-Z. (2021). Comprehensive review on facemask detection techniques in the context of covid-19. *IEEE Access*, 9:106839–106864. Disponível em 10.1109/ACCESS.2021.3100070.

OMS (2019). Coronavirus disease (COVID-19) pandemic. Disponível em https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Acesso em 18 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 18 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.

Prather, K. A., Wang, C. C., and Schooley, R. T. (2020). Reducing transmission of SARS-CoV-2. *Science*, 368(6498):1422–1424.

Purohit, A. (2020). Face mask dataset. Disponível em https://www.kaggle.com/aditya276/face-mask-dataset-yolo-format/version/2. Acesso em 18 de maio de 2022.

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.

Redmon, J. and Farhadi, A. (2017). Yolo9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, Estados Unidos.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. Disponível em https://arxiv.org/abs/1804.02767. Acesso em 18 de maio de 2022.

Shorfuzzaman, M., Hossain, M. S., and Alhamid, M. F. (2021). Towards the sustainable development of smart cities through mass video surveillance: A response to the COVID-19 pandemic. *Sustain. Cities Soc.*, 64(102582):102582.

Singh, S., Ahuja, U., Kumar, M., Kumar, K., and Sachdeva, M. (2021). Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. *Multimed. Tools Appl.*, 80(13):1–16.

Srivastava, N., Baxi, P., Ratho, R. K., and Saxena, S. K. (2020). Global trends in epidemiology of coronavirus disease 2019 (COVID-19). In *Medical Virology: From Pathogenesis to Disease Control*, pages 9–21. Springer Singapore.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. Disponível em 10.1109/CVPR.2001.990517. Acesso em 18 de maio de 2022.

Yang, G., Feng, W., Jin, J., Lei, Q., Li, X., Gui, G., and Wang, W. (2020). Face mask recognition system with YOLOV5 based on image recognition. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. IEEE.

Yang, S., Luo, P., Loy, C. C., and Tang, X. (2016). WIDER FACE: A face detection benchmark. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.