

Combate ao Covid19: Detecção em tempo real de indivíduos sem máscara em ambiente escolar por meio de Deep Learning

Elaine Barbosa Figueiredo, Eliana Pereira da Silva

Centro Universitário Campo Limpo Paulista – UNIFACCAMP

Campo Limpo Paulista – SP – Brasil

{efigueiredo13@gmail.com, eliana.pereiras@gmail.com}

Abstract. *Respiratory diseases such as COVID19 are transmitted by airways, especially in environments where it can be difficult to maintain physical distance between people and air circulation is not satisfactory, as is the case in the school environment. Health experts recommend the use of a protective face mask as a way to control the spread of this virus. In Brazil in particular, with the return of classroom format classes, a new concern arises regarding the use of protective masks by the school community, especially in the public education system, where spaces for collective use are not large and many individuals resist the use of this individual protection. Therefore, the study of a computational model using Deep Learning called YOLO was started, to detect people in the school environment who are not wearing a mask through the images of surveillance cameras. In the testing phase of this study, among all classifications, masked and unmasked subjects, which the model did 100%, were correct.*

Resumo. Doenças respiratórias como o COVID19 são transmitidas por vias aéreas, principalmente em ambientes onde pode ser difícil manter o distanciamento físico entre as pessoas e a circulação de ar não é satisfatória como é o caso do ambiente escolar. Especialistas da área da saúde recomendam o uso da máscara facial de proteção como forma de controlar a disseminação desse vírus. No Brasil em particular, com o retorno das aulas no formato presencial surge uma nova preocupação com relação ao uso de máscara de proteção por parte da comunidade escolar principalmente na rede pública de educação, onde os espaços de uso coletivo não são amplos e diversos indivíduos resistem ao uso dessa proteção individual. Diante disso, deu-se início ao estudo de um modelo computacional usando *Deep Learning* chamado *YOLO*, para detectar por meio das imagens das câmeras de vigilância as pessoas no ambiente escolar que não estão usando máscara. Na fase de testes desse estudo, dentre todas as classificações, indivíduos com máscara e sem máscara, que o modelo fez 100% estavam corretas.

1. Introdução

Doenças respiratórias propagadas por via aérea tais como: tuberculose, vírus influenza (gripe), síndrome respiratória aguda grave coronavírus 2 (SARS-CoV-2) causadora da

pandemia de COVID19, são transmitidas por meio de gotículas ou aerossóis expelidos quando um indivíduo que está contaminado com o vírus espirra, tosse, canta, grita ou até mesmo quando fala. Essas gotículas podem ser exaladas por outros indivíduos que estão próximos possibilitando assim, a contaminação do mesmo. A COVID-19 é uma infecção viral patogênica altamente contagiosa iniciada no mercado atacadista de frutos do mar de Wuhan na China em dezembro de 2019 e se espalhou rapidamente por todo o mundo devido à transmissão progressiva, causa infecções respiratórias, como resfriado comum, tosse seca, febre, dor de cabeça, dispneia, pneumonia e, finalmente, Síndrome Respiratória Aguda Grave (SARS) em humanos [MOHAPATRA et al. 2020].

Pesquisas recentes mostram a importância do uso da máscara facial no combate à disseminação de doenças respiratórias de modo particular a COVID19. De acordo com [LEUNG et al. 2020], o uso de máscara facial cirúrgica pode minimizar a transmissão dos vírus e consecutivamente reduzir o número de infectados. [KÄHLER et al. 2020] afirma que o uso de máscaras faciais sejam elas cirúrgicas ou artesanais pode oferecer 3 tipos diferentes de proteção: 1) evita tocar o rosto com as mãos; 2) reduz significativamente o risco de infecção nas proximidades de uma pessoa infectada e 3) reduz a possibilidade de inalação de gotículas contendo vírus que podem existir mesmo na ausência de pessoas infectadas. Para [SCHÜNEMANN et al. 2020], o uso de máscara de proteção individual não substitui outros meios de prevenção como: lavar as mãos, evitar aglomerações, redução no tempo de contato interpessoal, mas oferece benefícios por exemplo, em ambientes onde pode ser difícil manter o distanciamento físico, como é o caso do ambiente escolar. Além disso, diversos governantes estão orientando a população a fazer uso das máscaras faciais como forma de combater a contaminação e disseminação do vírus.

No Brasil em particular, com o retorno das aulas no formato presencial surge uma nova preocupação com relação ao uso de máscara de proteção por parte da comunidade escolar principalmente na rede pública de educação, onde os espaços de uso coletivo não são amplos e diversos indivíduos resistem ao uso dessa proteção individual.

Dado o exposto, deu-se início ao estudo de uma técnica usando Deep Learning para detectar por meio das imagens das câmeras de vigilância as pessoas no ambiente escolar que não estão usando máscara. Deep Learning é uma das formas mais conhecidas de se referir às DNNs (Redes Neurais Profundas). Uma técnica de DNN bastante poderosa e difundida são as CNNs (Redes Neurais Convolucionais) que são formadas por camada de entrada, camadas de convolução, camadas de pooling, camadas totalmente conectadas e camada de saída, e cada uma destas camadas possui uma função específica na propagação do sinal de entrada. Cada uma dessas camadas é constituída de vários planos celulares, e em cada plano celular, todos os neurônios têm as mesmas características, ou seja, realizam uma mesma função, porém em locais diferentes da imagem. Essa técnica de aprendizado profundo, permite que modelos computacionais compostos de várias camadas de processamento aprendam representações de dados com vários níveis de abstração [LeCun et al. 2015].

No desenvolvimento deste estudo foi usado o método computacional baseado em Deep Learning, chamado YOLO, You Only Look Once. A escolha do método YOLO deu-se pelo fato que o mesmo detecta um objeto sem ter que reclassificar. Em linhas gerais, o método YOLO usa uma rede neural profunda para reconhecer os objetos enquadrando o alvo que será detectado o que resulta em um sistema com uma velocidade maior que outros modelos computacionais utilizados em tarefas de detecção de objetos.

Nas próximas seções serão descritas as características técnicas do modelo computacional usado nesse estudo, iniciando pela Seção 2 onde se encontra uma revisão bibliográfica sobre o assunto. Na Seção 3 é descrita a metodologia de desenvolvimento do trabalho proposto. Na Seção 4 são apresentados os resultados e na Seção 5, as conclusões.

2. Trabalhos relacionados

As redes de aprendizado profundo (*Deep Learnig*) geram avanços na área de visão computacional, tais como a detecção de objetos de interesse numa imagem. Os pesquisadores têm desenvolvido trabalhos significativos como em [LOEY et al. 2021], que utiliza um modelo híbrido de aprendizado profundo e de máquina para detecção de máscara facial.

O modelo foi testado com três conjuntos de dados distintos O classificador SVM (*Sector Vector Machine* – Máquina de vetores de Suporte) atingiu 99.64% de precisão de teste usando o conjunto de dados RMFD (*Real-World Masked Face Dataset* – Conjunto de Dados Mascarado do Mundo Real). Já com o conjunto de dados SMFD (*Simulated Masked Face Dataset* - Conjunto de dados de rosto mascarado simulado), atingiu 99,49%, enquanto no conjunto de dados LFW (*Labeled Faces in the Wild* - Rostos marcados na natureza), atingiu 100% de precisão de teste. [LI et al. 2020] aplicou *YOLOv3*, após alguns ajustes ao problema de detecção de faces em ambientes complexos, de acordo com os resultados o desempenho não foi bom, em razão de que o *YOLOv3* tem reescrita intensiva dos rótulos no conjunto de dados impactando nos resultados [Liu et al. 2016]. Usou o detector de objetos SSD (*Single Shot Detector*) de disparo único para várias categorias. A característica principal desse modelo é o uso de saídas convolucionais de caixa delimitadora em várias escalas anexadas a mapas de características múltiplas no topo da rede. De acordo com os autores o SSD tem uma precisão melhor, mesmo com um tamanho de imagem de entrada menor. Em outra pesquisa [DEY et al. 2021] propôs um modelo de detecção de máscara chamado *Máscara MobileNet* que usa estrutura baseada em aprendizado profundo para treinar um classificador focado na detecção de máscaras faciais em dois conjuntos diferentes. Em testes realizados com 770 amostras de validação, a *Máscara MobileNet* atinge uma precisão de cerca de 93% enquanto que com 276 amostras de validação atinge uma precisão de aproximadamente 100%.

3 Metodologia

A proposta deste estudo é identificar em tempo real indivíduos sem máscara de proteção facial no ambiente escolar, por meio de imagens capturadas das câmeras de vigilância. Caso sejam detectados indivíduos sem a proteção facial, esta informação será enviada para a autoridade escolar competente para tomar as medidas necessárias. A Figura 1 exhibe o diagrama de blocos da estrutura do estudo proposto.

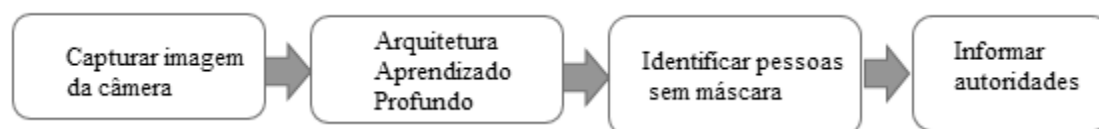


Figura 1 – Estrutura do estudo proposto

A. Estrutura Deep Learning

i) Rede em dois estágios: Nas arquiteturas de dois estágios o primeiro estágio responde pela extração das áreas de interesse (ROI), essa busca é feita utilizando algoritmo de busca seletiva para extrair as sub-regiões na imagem de entrada na qual está o objeto procurado. O segundo estágio responde pela classificação dos objetos detectados nas áreas de interesse. As previsões são realizadas em cada uma das regiões selecionadas aumentando drasticamente a velocidade de processamento da estrutura, todavia é capaz de atingir alta acurácia. R-CNN (*Region Based Convolutional Neural Networks* – Redes Neurais Convolucionais baseadas na região) e *Fast-RCNN* (*Fast Region-based Convolutional Network method* – Método rápido para redes neurais artificiais baseadas em região) são típicas arquiteturas em dois estágios.

ii) Rede em um estágio: YOLO integra operações de classificação e detecção em uma única etapa, o algoritmo olha apenas uma vez para a imagem, dividindo-a em diversas caixas delimitadoras (*bounding box*) e inferindo onde estão os objetos através de probabilidade. As caixas delimitadoras que possuem probabilidade acima de um determinado *threshold* são selecionadas, ou seja, nessas caixas há a possibilidade de existir um objeto reconhecido, [REDMON et al. 2016]. A Figura 2 representa a imagem de entrada e as etapas de delimitação e detecção respectivamente.

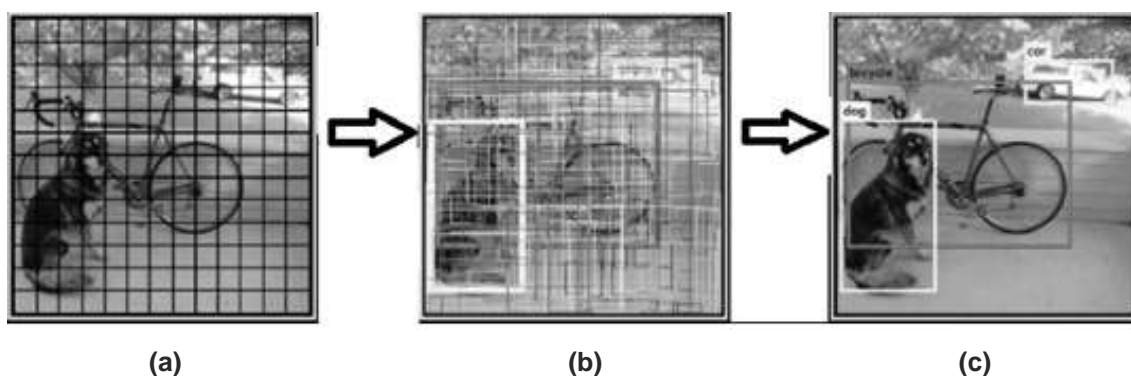


Figura 2 - (a) imagem de entrada (b) caixas delimitadoras dividindo a imagem (c) caixas delimitadoras com maior probabilidade

O modelo olha para a imagem inteira no momento do teste, de modo que suas previsões sejam informadas pelo contexto global da imagem. Ele também faz previsões com uma única avaliação de rede, ao contrário de sistemas com duas etapas. Isso o torna milhares de vezes mais rápido do que sistemas de duas etapas tais como *Fast R-CNN* e *R-CNN*. *Yolo* utiliza como *backbone* o *framework Darknet*, *Darknet* é uma rede neural profunda *open source* desenvolvida em linguagem *C* e *CUDA*, possuindo camadas convolucionais, filtros, funções de ativação e camadas totalmente conectadas, com suporte para uso de *GPU* (*Graphic Processing Unit* – Unidade Gráfica de Processamento). As camadas convolucionais iniciais da rede extraem atributos da imagem, enquanto que as camadas totalmente conectadas proveem as probabilidades e coordenadas de saída.

iii) *Google Colaboratory*: Tendo em vista que o foco deste estudo é a detecção de indivíduos sem máscara facial no ambiente escolar e que boa parte das escolas públicas no Brasil não dispõe de recursos computacionais robustos que suportem o alto custo computacional de uma rede neural profunda decidiu-se optar pelo ambiente *Google Colaboratory* para realizar os processos de treinamento e testes do modelo. O *Google Colaboratory* é uma ferramenta em nuvem formada de notebooks *Jupyter* que não requer

configuração executado no navegador. O serviço permite ainda escrever e executar códigos em *Python*, pois já vem com várias bibliotecas da linguagem pré-instaladas, o serviço permite inclusive habilitar a opção de utilizar Unidades de Processamento Gráfico (GPUs) para executar os códigos.

iv) *Conjunto de dados*: Para a realização das fases de treinamento e testes foi utilizado o conjunto de dados público disponível no Kaggle chamado *Face Mask Detection*, esse conjunto de dados contém 853 imagens e seus arquivos de anotação correspondentes, indicando se uma pessoa está usando uma máscara corretamente, incorretamente ou não. A Figura 3, mostra algumas imagens do conjunto de dados.



Figura 3 – Exemplo de imagens do conjunto de dados *Mask Detection*

v) *Rotulando os dados*: No contexto da detecção de objetos, rotular significa desenhar caixas delimitadoras em torno dos objetos que estamos interessados em detectar nas imagens e associá-los às classes. Os rótulos dos dados devem seguir o formato *YOLO*, que gera um arquivo de texto por imagem, com os seguintes dados: *object-id center_x center_y width height*, onde:

- *object-id*: representa o número correspondente à categoria do objeto listado no arquivo que contém os nomes das classes.

- *center_x and center_y*: representa o ponto central da caixa delimitadora. Mas eles são normalizados para ficarem dentro do intervalo 0 e 1, dividindo pela largura e altura da imagem o que torna mais fácil trabalhar com as mesmas, mesmo após dimensionar as imagens.

- *width and height*: representa a largura e a altura da caixa delimitadora.

Para rotular os dados foi usada a ferramenta chamada *RoboFlow*. A Figura 4, mostra algumas imagens rotuladas.



Figura 4- Imagens rotuladas

vi) *Informar as autoridades*: A proposta deste estudo é identificar, em tempo real, indivíduos dentro do ambiente escolar que não estão seguindo as orientações referente ao

uso de máscara facial como medida de prevenção no combate a disseminação do vírus da COVID19. O modelo computacional estudado destaca dentro de caixas delimitadoras a face dos indivíduos que não estão seguindo os protocolos sanitários. Quando um indivíduo for detectado sem essa proteção caberá ao diretor ou orientador educacional orientar e ou conscientizar o infrator, tais ações certamente ajudarão a limitar a disseminação da COVID19 no ambiente escolar.

4 Resultados

Para avaliação de desempenho considera-se a métrica *mean Average Precision* (mAP), esta métrica representa a precisão da identificação e classificação [EVERINGHAM et al. 2010]. Para a realização do treinamento do modelo personalizado dividiu-se o conjunto de dados em 70% para treinamento e 30% para testes. Ao todo, 6000 iterações foram executadas durante o processo de treinamento, o *framework Darknet* começa a mensurar a precisão a partir da milésima iteração, ao final desse processo a média de precisão foi 82,1%. A partir da iteração 2000 a perda média começou a baixar consideravelmente e ao final do processo de treinamento atingiu a marca de 0.62 com FPS (Frames por segundo) em torno de 144.

Objetivando avaliar o desempenho do modelo na detecção de indivíduos com e sem máscara, foram realizados testes em tempo real sobre as imagens capturadas por meio de uma *webcam* integrada ao notebook, este foi posicionado em local estratégico, em uma área com dimensões 5m x 3m e com boas condições de iluminação, que possibilitou capturar imagens das pessoas que acessavam esse espaço. Os testes foram divididos em três fases a saber:

a) Imagens de curta duração: Nesse teste o modelo foi posto à prova durante cinco minutos. Nas imagens capturadas haviam dois indivíduos;

b) Imagens de média duração: O modelo foi testado por um período de quinze minutos. Durante esse tempo foram capturadas imagens de três indivíduos;

b) Imagens de longa duração: O modelo foi posto à prova durante o período de uma hora. Nas imagens capturadas a quantidade de indivíduos sofreu diversas variações.

Em todas as imagens capturadas por meio da *webcam*, os indivíduos ora alternavam com e sem a máscara e também foram criadas várias situações de uso incorreto da máscara objetivando colocar à prova a capacidade do modelo na detecção de variações quanto ao uso incorreto da máscara. Neste sentido os indivíduos usaram a máscara apenas para cobrir a boca, pendurada na orelha, cobrindo nariz e olhos, abaixo do queixo, entre outras. Ao longo desses testes foi possível perceber que o modelo considera uso incorreto da máscara somente as ocorrências onde a mesma fica abaixo do queixo. A taxa de detecção foi significativa, conforme mostrado na Tabela 1.

Tabela 1 – Resultado dos testes

TEMPO	Com máscara	Sem máscara	Uso incorreto	Média FPS
Curta duração	100%	100%	97%	35.3
Média Duração	100%	100%	83%	
Longa Duração	100%	100%	79%	

5 Conclusão

Neste estudo foi considerado o modelo *YOLO* para realizar a tarefa de identificar em tempo real, por meio das imagens das câmeras de segurança, indivíduos dentro do ambiente escolar que não estão utilizando a máscara de proteção. Os resultados obtidos demonstram que essa técnica poderá ser aplicada com sucesso no controle da disseminação do vírus da COVID19 dentro do ambiente escolar, pois acredita-se que os indivíduos da comunidade escolar irão repensar na possibilidade de remover a máscara de proteção ao saber que poderão ser detectados.

Contudo, entende-se que para atingir precisão melhor na detecção de indivíduos usando máscara incorretamente é necessário aumentar o número de amostras no conjunto de dados que permita identificar as variações em relação ao uso incorreto da máscara. Como trabalhos futuros propõe-se:

- Extração dos resultados para análise dos dados objetivando a criação de um sistema de *BI* (*Business Intelligence* – Inteligência de Negócio) que permita obter diversas informações referente ao uso de máscara no ambiente escolar.
- Expandir o estudo para outros ambientes e contextos para observação do comportamento das pessoas em ambientes passíveis de aglomeração.
- Desenvolver um aplicativo *mobile* que faça o envio de notificações para os indivíduos que estejam violando as normas sanitárias, sejam alunos ou funcionários.

Referências

- Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C. (2020). Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* 26, 676–680
<https://doi.org/10.1038/s41591-020-0843-2>
- Kähler, C. J., Hain, R., (2020). Fundamental protective mechanisms of face masks against droplet infections, *Journal of Aerosol Science*, Volume 148
- Schünemann, H. J., Akl, E. A., Chou, R., Chu, D. K., Loeb, M., Lotfi, T., Mustafa, R. A., Neumann, I., Saxinger, L., Sultan, S., & Mertz, D. (2020). Use of facemasks during the COVID-19 pandemic. *The Lancet. Respiratory medicine*, 8(10), 954–955.
[https://doi.org/10.1016/S2213-2600\(20\)30352-0](https://doi.org/10.1016/S2213-2600(20)30352-0)
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015). <https://doi.org/10.1038/nature14539>
- Loey, M., Manogaran, G., Taha, M., & Khalifa, N. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement : journal of the International Measurement Confederation*, 167, 108288. <https://doi.org/10.1016/j.measurement.2020.108288>
- Li, C., Wang, R., Li, J. Fei, L., Face detection based on YOLOv3, in: *Recent Trends in Intelligent Computing, Communication and Devices*, Singapore, 2020, pp. 277–284, doi: 10.1007/978-981-13-9406-5_34.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C., (2016) SSD: Detector MultiBox de disparo único. Em: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision - ECCV 2016*. ECCV 2016. Lecture Notes in

Computer Science, vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2

Dey, S.K., Howlader, A., Deb, C. (2021) MobileNet Mask: A Multi-phase Face Mask Detection Model to Prevent Person-To-Person Transmission of SARS-CoV-2. In: Kaiser M.S., Bandyopadhyay A., Mahmud M., Ray K. (eds) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing, vol 1309. Springer, Singapore. https://doi.org/10.1007/978-981-33-4673-4_49

Redmon, j., Divvala, S., Girshick, R., and Farhadi, A., (2016). You Only Look Once: Unified, Real-Time Object Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

Everingham, M., Van Gool, L., Williams, C.K.I. (2010). The PASCAL Visual Object Classes (VOC) Challenge. *Int J Comput Vis* **88**, 303–338 <https://doi.org/10.1007/s11263-009-0275-4>

Mohapatra, R. K., Pintilie, L., Kandi, V., et al. (2020). The recent challenges of highly contagious COVID-19 , causing respiratory infections : Symptoms , diagnosis , transmission , possible vaccines , animal models , and immunotherapy. n. June, p. 1187–1208