

Classificação do uso de Máscaras de Proteção utilizando CNN para detecção do Risco de Contaminação do Covid-19

1st Passamani C. G.
Departamento de Informática
Universidade Federal do Espírito Santo
Vitória, Brasil
gobbo.cezar@gmail.com

2nd Melo R. C.
Departamento de Informática
Universidade Federal do Espírito Santo
Vitória, Brasil
rod.cdr@gmail.com

Abstract—Monitorar manualmente se indivíduos estão usando máscaras de proteção contra o Covid-19 corretamente em áreas públicas e lotadas é uma tarefa difícil. Este artigo aborda uma maneira simplificada de obter a detecção se o indivíduo está ou não está usando máscara facial em um sistema de tempo real. A máscara é extraída de rostos em tempo real em público e é alimentada como uma entrada na rede neural convolucional (CNN).

Keywords—mask, social distancing, CNN, Covid-19

1. INTRODUÇÃO

A pandemia da covid é um problema enfrentado por todos os países, no Brasil superamos no início de outubro de 2021 mais de 600 mil mortes [1], e com menos de 100 milhões de brasileiros com a segunda dose ou única da vacina contra a covid aplicadas [2]. Com a vacinação os números de novas infecções têm diminuído, assim com a quantidade de internações, entretanto, outras medidas de segurança são necessárias para que obtenhamos êxito no combate ao coronavírus. Duas medidas que são consideradas efetivas pela OMS são o uso de máscaras e o distanciamento social [3]. Nesse artigo detectamos o uso de máscara utilizando a rede neural de aprendizado profundo (especificamente o modelo YOLO-darknet). Na visão computacional, métodos de reconhecimento de imagem tipicamente dividem este processo em duas etapas: a detecção dos objetos e a subsequente classificação deles. A YOLO tem a vantagem de unificar o processo de detecção de objetos e a classificação dos mesmos, precisando “olhar” a imagem apenas uma vez, daí a origem do nome – *You Only Look Once*.

2. MÉTODO

O diagrama mostra passo a passo o processo utilizado, mostrado na figura 1.

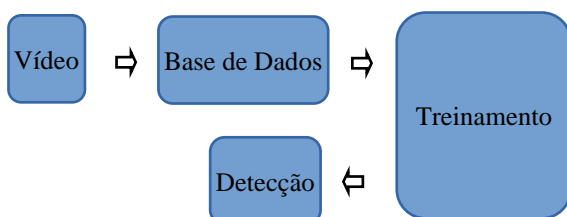


Figura 1. Diagrama de blocos do processo.

1. Vídeos Obtidos

Vídeos gravados com rostos são os dados de entrada para esse estudo. É importante que haja variação nos vídeos, como por exemplo de iluminação, distância entre a câmera e o rosto. Esses vídeos são o ponto de partida para a criação de uma base de dados usando extração de características.

2. Base de Dados

Para a detecção de máscaras faciais, não conseguimos encontrar pesos pré-treinados para a rede neural YOLO. Sendo um problema no campo de detecção de objetos recente, é natural a dificuldade de encontrar uma base de dados consolidada, então foi necessário compilar nossa própria base de dados com imagens e treinar a rede de detecção de máscaras do zero.

Apesar de imagens de rostos mascarados estarem disponíveis de forma abundante na internet, não são muitas que trazem pessoas utilizando a proteção longe da fonte da imagem, e nosso sistema prevê o uso de câmeras em lugares abertos e ao ar livre, portanto necessitávamos de imagens com estas características para realizar o treinamento de uma rede adequada.

Consideramos então o problema de detecção de máscaras faciais como uma subclasse do problema de detecção de rostos, que já é bem pesquisado e apresenta boas soluções. Nesse caso, deveríamos produzir um modelo de detecção de rostos que produz duas classificações destes objetos detectados: os rostos sem máscara, e o rosto com máscara.

Para o treinamento da detecção de rostos, existem boas bases de imagens. Optamos pela WIDER FACE [4] com 393.703 imagens de rostos. Esta base não contém especificamente uma classe de rostos mascarados, mas foi possível selecionar algumas imagens da base de rostos com imagens de pessoas sem máscara para montar nossa base de treinamento. Utilizamos, portanto, um editor de imagens para manualmente inserir máscaras nos rostos de parte das imagens da WIDER FACE de dados, como mostra o antes e depois na figura 2.



Figura 2. Antes e depois da inserção de manual de máscaras na criação da base de dados.

Outra solução que encontramos foi capturar vídeos abertos de câmeras ao vivo em locais públicos, onde era possível ver muitas pessoas transitando utilizando máscaras. Separamos imagens destes vídeos e fizemos a marcação manual dos mesmos, como mostra a figura 3, essa marcação foi feita utilizando o programa Labellmg disponível no link: <https://github.com/tzutalin/labelImg>.

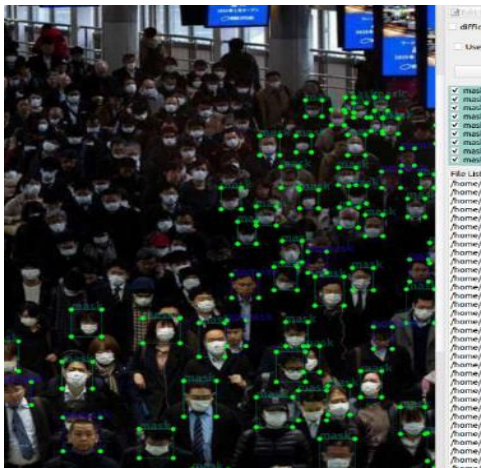


Figura 3. Marcação manual de novos objetos para a rede neural.

Repetimos também este procedimento em cima de vídeos aleatórios capturados do YouTube que realizaram noticiários sobre a pandemia e conseguiam registrar cidades e ruas durante a pandemia, como mostra a figura 4.



Figura 4. Pessoas utilizando máscaras em vídeos do YouTube

Ao final juntamos 24.923 imagens para treinar a rede. As imagens foram separadas em 2 classes, “com máscara” e “sem máscara”.

As imagens do dataset podem ser encontradas no link abaixo:

https://drive.google.com/file/d/1IXgEWbvUEb_xl0CCEmz24gK7Tm8VFqN/view?usp=sharing

3. Treinamento

O treinamento foi realizado em um notebook com uma placa de vídeo RTX 2060, utilizando 4984 (20%) imagens para validação e 19939 (80%) imagens para treino. Na figura 5, é mostrado o resultado do treinamento da rede, com as informações do loss e do mean average precision (mAP).

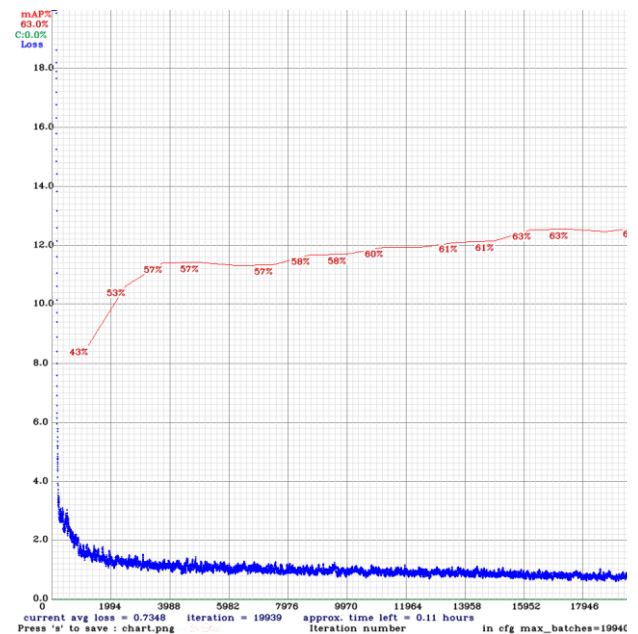


Figura 5. Gráfico do resultado do treino no Darknet.

A tabela abaixo mostra os hiperparâmetros utilizados, a acurácia e a perda.

Batch	64
Image Width	256
Image Height	256
Learning Rate	0,001
Max Batches	19940
Accuracy	63%
Loss	0,7348
Validation Images	4984 (20%)
Training Images	19939 (80%)

Tabela 1. Parâmetros e resultados do treino.

Assim com a rede treinada, fomos capazes de identificar nos vídeos quais rostos de pessoas utilizam ou não máscara.

3. ALGORÍTMO

O algoritmo de detecção do risco de contaminação é simples e foi feito da seguinte forma: dada uma imagem, o algoritmo infere a imagem na rede neural de detecção do uso de máscaras. A rede retorna quantos rostos estão utilizando máscaras de proteção e quantos não estão. Se na imagem tiver 0 ou 1 rosto com ou sem máscara, é retornado risco baixo (low) de contaminação, se tiver entre 2 e 4 rostos

sem máscaras, é retornado risco médio (medium), se tiver entre 5 e 7 rostos sem máscara, é retornado risco alto (high), se tiver mais de 8 rostos sem máscara, é retornado risco altíssimo (very high). Exemplos de imagens com esses riscos estão nas figuras 6, 7, 8 e 9.

4. EXPERIMENTO E RESULTADOS

O resultado das execuções da rede em vídeos encontrados na internet foi bem satisfatório. A figura 6 mostra um exemplo do resultado da detecção da rede, onde existem 8 pessoas na praia sem máscara, fazendo com que o risco de contaminação seja muito alto, como é mostrado no canto superior esquerdo. O quadrado amarelo no rosto das pessoas é o resultado da detecção facial, tendo ao lado o label resultante da classificação da rede, podendo ser “nomask” para detecção de rosto sem máscara, e “mask” para detecção de rosto com máscara. Quando o resultado da classificação é “nomask” o label fica vermelho, e quando é “mask” o label fica verde.



Figura 6. Pessoas na praia aglomeradas sem máscara.

O experimento no vídeo da figura 7 foi feito a partir de um vídeo de um restaurante, onde aproximadamente 24 pessoas estão sem máscara. O resultado do risco de contaminação foi muito alto pois todas as pessoas estavam sem máscara.



Figura 7. Pessoas sem máscaras em um restaurante.

O experimento no vídeo da figura 8 foi feito a partir de um vídeo na escadaria de um prédio, onde 5 homens não estão usando máscara de proteção gerando um alto risco de contaminação.



Figura 8. Pessoas sem máscaras em uma escadaria.

A imagem da figura 9 veio do resultado da detecção em um vídeo de um metrô na Coreia, onde todas as pessoas estão usando máscara de proteção, fazendo com que o resultado do risco de contaminação seja baixo.



Figura 9. Pessoas com máscaras em um metrô na Coreia.

Os vídeos utilizados nos experimentos podem ser encontrados através do link:

https://drive.google.com/file/d/1LsoZaNLjhlHnvGHhpJL9P2_eZ-qiQUm2/view?usp=sharing

5. REFERÊNCIAS

- [1] Coronavírus Brasil. [online]. Disponível em: <https://covid.saude.gov.br/>. [Acessado em 12 de Outubro de 2021].
- [2] Gov.br. 2021. Vacinação contra a Covid-19 no Brasil - #PÁTRIAVACINADA. [online] Disponível em: <https://www.gov.br/saude/pt-br/vacinacao/> [Acessado em 12 de Outubro de 2021].
- [3] Who.int. 2021. When and how to use masks. [online] Disponível em: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks> [Acessado em 14 de Outubro de 2021].
- [4] S. Yang, P. Luo, C. C. Loy, and X. Tang, “Wider face: A face detection benchmark,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.