YOLOV4 APLICADO À IDENTIFICAÇÃO DO USO CORRETO DE MÁSCARAS FACIAIS DIANTE DE SURTOS VIRAIS

Francisco Alan de França Pereira*
Mário Wedney de Lima Moreira**

RESUMO

Em razão do grande número de contaminações e óbitos nos últimos anos decorrentes de crises virais, como a causada pelo novo coronavírus SARS-CoV-2 (COVID-19), o uso de máscaras faciais tem se mostrado eficaz na contenção destas doenças. Em razão disto, observa-se que as técnicas de reconhecimento facial apresentam uma solução eficaz na fiscalização do uso de máscaras em ambientes onde o risco de contaminação é maior, como escolas, hospitais e supermercados. A fim de criar um modelo preditivo capaz de identificar se as pessoas estão usando efetivamente máscaras faciais, este artigo apresenta uma solução baseada em redes neurais convolucionais que usa o modelo preditivo YOLOv4, que é capaz de identificar em tempo real, a partir de imagens, se as pessoas em determinado ambiente estão com ou sem máscara. Para tal, a partir de um conjunto de imagens coletadas, rotuladas e submetidas ao *framework* Darknet, o modelo foi treinado e testado. Como resultado, verifica-se que esta proposta obteve uma *mean average precision* (mAP) de 99,36%, precisão de 97% e recall de 99%. Estes resultados constatam que o YOLOv4 mostra-se como uma solução eficaz quanto a fiscalização do efetivo uso de máscaras em crises virais.

Palavras-chave: Redes neurais convolucionais. Reconhecimento de imagens. Síndromes espiratórias agudas graves.

ABSTRACT

Due to the large number of contamination and deaths in recent years resulting from viral crises, such as the one caused by the new coronavirus SARS-CoV-2 (COVID-19), the use of face masks has proven to be effective in containing these diseases. As a result, it is observed that facial recognition techniques present an effective solution for monitoring the use of masks in environments where the contamination risk is greater, such as schools, hospitals, and supermarkets. To develop

* Graduando em Ciência da Computação, Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Aracati, CE, Brasil. E-mail: francisco.alan.franca00@aluno.ifce.edu.br.
** Doutor em Engenharia Informática, docente do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Aracati, CE, Brasil. E-mail: mario.wedney@ifce.edu.br.

a predictive model capable of identifying whether people are effectively using face masks, this paper presents a solution based on convolutional neural networks that use the YOLOv4 predictive model, which is capable of identifying in real-time, from images, whether people in a given environment are wearing or not wearing a mask. For this, from a set of images, collected, labeled, and submitted to framework Darknet, the model was trained and tested. As a result, it appears that this proposal obtained a mean average precision (mAP) of 99.36%, an precision of 97%, and a recall of 99%. These results show that YOLOv4 is an effective solution for monitoring the effective use of masks in viral crises.

Keywords: Convolutional neural networks. Image recognition. Severe acute respiratory syndromes.

1 INTRODUÇÃO

Em dezembro de 2019 deu-se início a uma pandemia causada pelo novo coronavírus nomeado de SARS-CoV-2 (COVID-19). Os primeiros casos de COVID-19 foram identificados em Wuhan, cidade na província de Hubei, no centro da China. Observando o crescimento e disseminação do vírus entre os seres humanos, a Organização Mundial da Saúde (OMS) declarou o coronavírus como uma pandemia mundial em março de 2020. Segundo (MOHAPATRA et al., 2020), a COVID-19 é uma infecção viral patogênica altamente contagiosa que se espalhou rapidamente por todo o mundo devido à sua fácil transmissão. Esta causa diversas infecções respiratórias, *e.g.*, resfriado comum, tosse seca, febre, dor de cabeça, dispneia, pneumonia e, finalmente, a síndrome respiratória aguda grave (SARS, em inglês) em humanos.

Pesquisas indicam que usar máscara, manter o distanciamento social e fazer a higiene das mãos ajudam a retardar a propagação do coronavírus (WHO, 2020). Kahler *et al.* afirma que máscaras faciais são instrumentos eficazes no combate à propagação do coronavírus, assim como outras doenças respiratórias, pois estas diminuem a propagação e a inalação de gotículas de saliva, reduzindo a disseminação do vírus (KÄHLER; HAIN, 2020). Este acessório evita o toque das mãos no rosto de uma pessoa. Como resultado, vários departamentos de saúde em todo o mundo começaram a emitir diretrizes para seus respectivos países.

No Brasil alguns fatores contribuíram para diminuição do uso de máscaras. O descuido com a saúde e distribuição de *fake news* contribuíram para reduzir a adesão a este equipamento de proteção. Nesse cenário, uma ferramenta de reconhecimento facial para identificar o uso efetivo e correto de máscaras em ambiente de intenso fluxo de pessoas, como escolas, pode ser um fator essencial para o controle de surtos epidêmicos de doenças transmitidas por vias respiratórias.

Diante do exposto, uma solução baseada em *deep learning* auxiliaria no controle de acesso a ambientes onde o uso da máscara é obrigatório. (GOODFELLOW; BENGIO; COURVILLE, 2016) descrevem esta técnica como um tipo específico de aprendizado de máquina que faz uso de um algoritmo de aprendizado ajustando diferentes dados de treinamento com o objetivo de

encontrar padrões que os generalizem para classificar novos dados submetidos .

Entre as possíveis soluções na área de reconhecimento facial, este estudo decidiu utilizar uma solução chamada de YOLO (*you only look once*, em inglês). Esta abordagem consiste na utilização de redes neurais convolucionais (RNCs) da área de *deep learning*. A decisão foi tomada em razão de seu potencial para utilização em aplicações comerciais, sua versatilidade de implementação disponível em diversas plataformas e sua alta velocidade de detecção, mesmo sem ter à disposição um *hardware* muito robusto.

Esse artigo visa criar e avaliar resultados de um modelo preditivo baseado em *deep learning* usando o YOLOv4. A partir de um *dataset* de treinamento, busca-se classificar de forma supervisionada as entradas de imagens em duas classes, a saber, "com máscara" e "sem máscara". Ao monitorar o efetivo uso da máscara, objetiva-se ajudar a conter a propagação do vírus durante surtos. O modelo proposto oferece uma solução que não depende de grande processamento computacional e pode ser aplicado a circuitos de televisão e integradas a portas eletrônicas de acesso a ambientes restritos. A triagem de quantidades maiores de pessoas é possível, portanto, pode esta abordagem pode ser usada em lugares de grande circulação, *e.g.*, bancos, mercados, hospitais, escolas, faculdades, entre outros.

Em relação à estrutura do artigo, na Seção 2 está definida a fundamentação teórica apresentando uma contextualização sobre conceitos relativos a inteligência artificial (IA) e sua ralação com o projeto desenvolvido. Na Seção 3 são abordados os trabalhos relacionados em que são analisados estudos publicados por outros autores sobre a temática. Na Seção 4, apresenta-se a metodologia adotada para a aplicação proposta. Na Seção 5 são apresentadas dados a respeito do *dataset* utilizado, assim como os resultados obtidos após o treinamento da rede neural. Finalmente, na Seção 6 são expostas as conclusões obtidas.

2 FUNDAMENTAÇÃO TEORICA

2.1 Inteligência artificial

Segundo (FERNANDES, 2004), a IA provém do latim que significa *inter* (entre) e *legere* (escolher), *i.e.*, este conceito está relacionado à atribuição de um algoritmo a uma máquina, dando capacidade para esta de escolher entre uma opção ou outra. A inteligência também é uma forma de resolver problemas, realizando tarefas. Então, considera-se a IA um tipo de inteligência produzida computacionalmente para beneficiar as máquinas, a fim de atribuir habilidades que simulam o comportamento humano. Pode-se citar, taxtite.g., compreensão da linguagem natural, reconhecimento de imagens e objetos, tomada de decisão, automação, entre outros.

No que diz respeito às máquinas, pode-se dizer, de forma muito ampla, que uma máquina aprende sempre que muda sua estrutura, programa ou base de dados (com base em suas entradas ou em resposta a informações externas) de tal maneira que seu desempenho melhora conforme seu algoritmo é treinado. Algumas dessas mudanças, como quando busca-se realizar reconhecimento facial de um registro em uma base de dados, caem confortavelmente dentro da província de

outras disciplinas e possivelmente melhor compreendidas por ser chamada de aprendizagem, quando, por exemplo, uma máquina de reconhecimento facial melhora seu desempenho após a visualização de várias amostras de imagens de uma pessoa. A partir daí, nota-se, nesse caso, que a máquina aprendeu (NILSSON, 1996).

Existem dois principais paradigmas de aprendizado de máquina. Estes são o aprendizado supervisionado e o não supervisionado, a saber:

- **Aprendizagem supervisionado**: Os algoritmos de aprendizagem supervisionada relacionam uma saída com uma entrada com base em dados já rotulados (dados já conhecidos). Neste caso, o usuário alimenta o algoritmo com um conjunto de entradas vinculadas a uma saída específica que por sua vez forma uma classe. Isso acontece, por exemplo, quando submetemos um conjunto de fotos de maçãs estes por sua vez formam uma classe e o algoritmo passa a classificar as futuras entradas baseadas nos dados pelo qual foi treinado (HAN MICHELINE KAMBER, 2011);

- **aprendizagem não-supervisionada**: No caso dos algoritmos deste tipo de aprendizagem, não acontece a rotulação nem a vinculação às saídas específicas. Através da submissão de um grande número de dados, o algoritmo estabelece padrões e similaridades entre estes, permitindo identificar grupos de itens similares ou similaridade entre novos itens submetidos, que podem ser classificados conforme sua tipicidade ou atipicidade (HAN MICHELINE KAMBER, 2011).

2.2 *Deep learning*

Deep learning é uma forma de aprendizado de máquina que possui grande poder e flexibilidade, pois permite que os computadores aprendam com a experiência e entendam o mundo em termos de uma hierarquia de conceitos. Como o computador reúne conhecimento baseado na experiência, não há necessidade de um operador de computador humano especificar formalmente todo o conhecimento de que o computador precisa. A hierarquia de conceitos permite que o computador aprenda conceitos complicados construindo-os a partir de conceitos mais simples. Um gráfico dessas hierarquias teria muitas camadas de profundidade (GOODFELLOW; BENGIO; COURVILLE, 2016).

Jiang *et al.* (2022) acrescentam que esta abordagem é baseada em múltiplas camadas de neurônios, que tratam matematicamente os dados. Estes são capazes de realizar tarefas como compreender a fala humana e reconhecer objetos visualmente (JIANG et al., 2022). A informação é processada através de suas camadas, com a saída da camada anterior repassando a entrada para a próxima camada. A primeira camada em uma rede é chamada de camada de entrada, enquanto a última é chamada de camada de saída. Todas as camadas entre a camada de entrada e a de saída são referidas como camadas ocultas. Cada camada é normalmente um algoritmo simples e padronizado possuindo um tipo de função de ativação.

Além disso, para (WANG; DENG, 2021), os modelos de *deep learning* podem resolver problemas mais complexos de forma rápida e eficaz, devido aos modelos mais complexos aplicados. Estes modelos podem aumentar consideravelmente a precisão da classificação ou até mesmo reduzir o erro nos problemas de regressão, desde que existam conjuntos de dados adequados com grande quantidade de detalhes. Esses fatores permitem às estruturas de *deep learning* serem consideravelmente flexíveis e adaptáveis para tratar uma ampla variedade de desafios altamente complexos. Embora este método seja aplicável para resolver diversos problemas, este mostra-se especialmente promissor para resolver problemas de identificação de faces.

Alguns dos mais comuns tipos de *deep learning* são de aprendizado supervisionado, não supervisionado, por reforço, redes neurais recorrentes (RNNs, do inglês), RNCs, redes de transformadores (*transformers*, do inglês), entre outros.

2.3 Redes Neurais Convolucionais

As RNCs são um tipo especializado de rede neural projetada especificamente para processar dados de formato estruturado, como imagens e sinais de áudio. Elas são amplamente usadas em tarefas de visão computacional, como reconhecimento de objetos, classificação de imagens e segmentação de objetos.

Segundo (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) uma RNC é composta por camadas de neurônios chamadas camadas convolucionais. Cada neurônio em uma camada convolucional está conectado a uma pequena região localizada na entrada (imagem). Essa conexão local é chamada de campo receptivo. Os neurônios aplicam operações de convolução nesses campos receptivos para extrair características específicas das imagens, como bordas, texturas e padrões visuais. Os autores acrescentam que, além das camadas convolucionais, as RNCs também podem incluir camadas de *pooling*, que reduzem a dimensionalidade das características extraídas pela seleção das características mais relevantes. Isso ajuda a reduzir a quantidade de parâmetros e a tornar o modelo mais eficiente. Em seguida, podem ser adicionadas camadas totalmente conectadas, semelhantes às de redes neurais tradicionais, que realizam a classificação final com base nas características aprendidas.

Durante o treinamento, as RNCs ajustam os pesos e os viéses dos neurônios para otimizar o desempenho do modelo. Isso é feito por meio do algoritmo de retropropagação, que propaga o erro calculado na camada de saída de volta para as camadas anteriores, atualizando os pesos ao longo do caminho (GOODFELLOW; BENGIO; COURVILLE, 2016).

3 TRABALHOS RELACIONADOS

No campo da IA aplicada ao reconhecimento facial são muitos os modelos baseados em *deep learning* utilizados entre os trabalhos desenvolvidos para identificação do efetivo uso de máscara.

Em razão da pandemia da COVID-19, em junho de 2020 não havia nenhum medicamento com eficácia comprovada para esta síndrome respiratória, visando seguir as recomendações

da Organização Mundial da Saúde (OMS). A solução proposta por (YADAV, 2020) visou fazer controle do uso de máscaras e o distanciamento social em lugares públicos que, quando descumprido, um aviso seria emitido às autoridades competentes. A aplicação proposta fez uso de uma *single shot detector multiBox* (SSD) para detecção de objetos em tempo real. O modelo VGG-16 é pré-treinado no ImageNet como seu modelo básico para extrair recursos de imagem úteis combinado com uma rede neural leve MobileNetV2 como técnica de aprendizagem de transferência para atingir o equilíbrio das limitações de recursos e precisão de reconhecimento e OpenCV. O modelo foi proposto para funcionar em uma Raspberry pi 4. A base de dados relatada é formada por 3.165 imagens usadas para identificar duas possíveis condições, a saber, a de uso correto da máscara e do uso incorreto da máscara. Também é feito o cálculo da distância euclidiana entre os pontos para determinar a distância entre as pessoas. Foi relatado que o modelo aplicado detecta o distanciamento social e as máscaras com pontuação de precisão de 0,917, confiança de 0,7, *recall* de 0,91 e com FPS igual a 28,07.

Impulsionado pelo avanço da pandemia de COVID-19 e toda problemática decorrente do controle de acesso de pessoas usando usando adequadamente a máscara em ambientes públicos e privados, (LIU; REN, 2021) observaram que muitas das soluções existentes para identificação do uso de máscaras eram desenvolvidas usando Mask-R-CNN com ResNet como *backbone* por ser um modelo de *deep learning*. Esta proposta sugeriu a substituição deste modelo por um modelo YOLO. Para tanto foi realizado um projeto que visa comparar o desempenho do modelo YOLOv3 usando Darknet e o modelo Mask-R-CNN com ResNet, ambos usaram o conjunto de treinamento disponibilizado pela plataforma Kaggle composto por 853 imagens. Após ajustes e visando a melhor otimização do *dataset* de imagens, o modelo YOLOv3 obteve uma acurácia de 0,932. Desempenho bem similar ao Mask-R-CNN. Este resultado valida que a família YOLO possui grande capacidade em relação ao estado da arte das tecnologias atuais de reconhecimento de objetos. Pode-se ainda salientar que seu tempo de execução é bem menor, o que a torna muito relevante no desenvolvimento de aplicações de detecção em tempo real.

Figueiredo e Silva (2021) desenvolveram um projeto visando testar o modelo YOLO para identificação do uso de máscaras em ambiente escolar levando em conta o retornos às aulas, em especial, na rede pública (FIGUEIREDO; SILVA, 2021). Este estudo visou medir a acurácia do modelo YOLOv4 para a identificação de máscaras em alunos. Um dataset com cerca de 853 imagens foi usado, no qual não foi mencionado a quantidade de objetos de cada tipo continha na base de treinamento e nem suas proporções. Deste número de imagens 70% foram usadas para treinar o algoritmo e 30% para realizar os testes ao final de 6.000 épocas de treinamento (iterações do algoritmo). A precisão geral do algoritmo ficou em cerca de 82,1%. O teste do modelo foi realizado por meio de *webcam* durante três intervalos de tempo 5, 15 e 60 minutos, respectivamente. Neste teste, o algoritmo foi capaz de identificar o uso e a ausência de máscaras em 100% dos casos. No entanto, o uso incorreto teve desempenho menor, em torno de 79% no pior cenário.

4 PROPOSTA

Este estudo propõe o desenvolvimento de um modelo de reconhecimento facial utilizando a arquitetura YOLOv4. Este modelo foi aplicado em um controle de acesso a ambientes onde o risco de contaminação viral por doenças infecto-contagiosas transmitidas pelo ar é maior. O YOLOv4 é um dos mais avançados modelos de detecção de objetos em tempo real e possui uma estrutura neural profunda capaz de extrair características complexas das imagens.

Ao adaptar essa arquitetura para o reconhecimento facial, pode-se treinar a IA em um amplo conjunto de dados de rostos e ensiná-la a identificar e classificar faces com alta precisão. Além disso, ao utilizar o YOLOv4, pode-se obter um processamento em tempo real, permitindo a aplicação do modelo em diversos cenários, como em ambientes de vigilância, impedindo o transito de pessoas desprotegidas (sem máscara). Com esta proposta, espera-se contribuir para o avanço da tecnologia de reconhecimento facial e possibilitar o desenvolvimento de sistemas mais eficientes e confiáveis no combate a propagações virais por meio de IA e *deep learning*.

4.1 YOLOv4

O modelo preditivo YOLOv4 é uma versão avançada de um sistema de detecção de objetos em tempo real. Ele foi desenvolvido para realizar a detecção e classificação de objetos em uma cena de forma rápida e precisa. O YOLOv4 apresenta várias características e vantagens que o tornam uma escolha popular em aplicações de visão computacional.

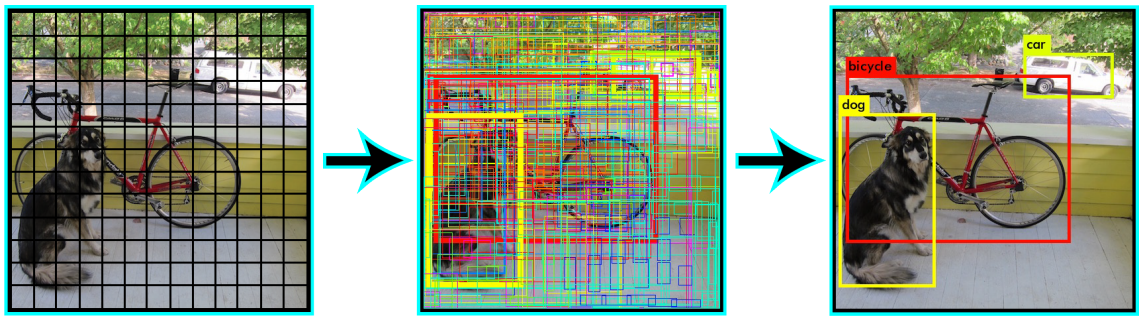
Uma das principais características do YOLOv4 é a sua capacidade de realizar detecção de objetos em tempo real. Isso significa que o modelo é capaz de analisar cada quadro de um vídeo ou imagem em tempo real, identificar objetos presentes e fornecer as coordenadas de suas caixas delimitadoras juntamente com as classes correspondentes. Essa capacidade é extremamente valiosa em aplicações em que a velocidade de resposta é crucial, como em sistemas de segurança, veículos autônomos e monitoramento de tráfego.

Além da velocidade, o YOLOv4 também se destaca pela sua precisão. O modelo utiliza uma arquitetura de rede neural profunda, que é treinada em grandes conjuntos de dados anotados. Isso permite que o modelo aprenda a reconhecer padrões complexos e faça previsões precisas sobre a presença e a classe dos objetos na cena. A precisão do YOLOv4 é particularmente impressionante em comparação com outros modelos de detecção de objetos em tempo real, tornando-o uma escolha popular para aplicações de missão crítica.

Além disso, o YOLOv4 também se beneficia de técnicas avançadas de visão computacional, como a fusão de recursos multi-escala e o uso de atenção espacial. Essas técnicas ajudam o modelo a extrair informações contextuais e espaciais das imagens, permitindo que este classifique de forma mais precisa sobre a localização e classe dos objetos. Essas melhorias levam a um desempenho aprimorado do modelo em termos de taxa de detecção e precisão. A Figura 1 apresenta um exemplo desta técnica.

O YOLOv4 usa *bounding boxes* para identificar e localizar os objetos em uma imagem. Estas caixas delimitadoras, também conhecida como *bbox*, são uma técnica utilizada em detecção

Figura 1 – Exemplo do uso do YOLOv4 para a detecção de objetos em imagens.



Fonte: IA Expert Academy.

de objetos para delimitar visualmente a localização destes em uma imagem. Esta consiste em um retângulo que envolve o objeto detectado, especificando suas coordenadas (x, y) do canto superior esquerdo e suas dimensões de largura e altura. Durante o processo de detecção, o modelo gera uma série de *bounding boxes* que representam as possíveis localizações dos objetos na imagem. Estas são geradas em diferentes escalas e posições, permitindo que o modelo localize objetos de tamanhos variados.

Cada *bounding box* gerada pelo YOLOv4 é associada a uma pontuação de confiança, que indica o quão provável é a presença de um objeto naquela região. Essa pontuação é calculada com base em várias características, como a resposta da RNC às características da imagem e a probabilidade de pertencer a uma classe específica. No entanto, é importante observar que o YOLOv4 pode gerar várias *bounding boxes* para um único objeto em diferentes escalas e posições. Para eliminar a redundância e selecionar estas caixas delimitadoras mais relevantes, o modelo utiliza uma técnica chamada supressão não máxima, *non-maximum suppression* (NMS), em inglês. A NMS analisa as sobreposições entre as *bounding boxes*, calcula uma pontuação de sobreposição e mantém apenas as caixas com pontuações mais altas, descartando as redundantes.

Dessa forma, o YOLOv4 usa *bounding boxes* para localizar e delimitar os objetos detectados na imagem. Estas são geradas em diferentes escalas e posições, e sua pontuação de confiança é utilizada para selecionar as detecções mais relevantes. Através do uso de *bounding boxes* e técnicas de pós-processamento, o YOLOv4 é capaz de realizar detecção de objetos em tempo real com alta precisão.

4.2 Matriz de confusão

Uma matriz de confusão é uma ferramenta usada para avaliar o desempenho de um modelo de classificação ou de um sistema de previsão. Esta mostra a relação entre as classificações reais e as classificações previstas feitas pelo modelo. A Figura 2 mostra esta relação.

- **Verdadeiros positivos (VP):** são os casos em que a classe real é positiva e o modelo classificou corretamente como positiva.
- **Falsos positivos (FP):** são os casos em que a classe real é negativa, mas o modelo erroneamente classificou como positiva.

Figura 2 – Matriz de confusão.

Matriz de Confusão		Detectado	
		Sim	Não
Real	Sim	*Verdadeiro Positivo (VP)*	*Falso Negativo (FN)*
	Não	*Falso Positivo (FP)*	*Verdadeiro Negativo (VN)*

Fonte: Imagem própria dos autores.

- **Verdadeiros negativos (VN):** são os casos em que a classe real é negativa e o modelo classificou corretamente como negativa.
- **Falsos negativos (FN):** são os casos em que a classe real é positiva, mas o modelo erroneamente classificou como negativa.

4.3 Metricas de avaliação

Para analizar o desenpenho do modelo preditivo iremos aplicar as seguintes metricas de avaliação.

4.3.1 Average Precision (AP):

Esta métrica avalia a precisão média das detecções em diferentes categorias de objetos. É calculada como a área sob a curva de precisão-*recall* (AP-R).

- A precisão é a proporção de verdadeiros positivos em relação ao número total de detecções. Ela indica a precisão do modelo em identificar corretamente os objetos. O *recall*, é a proporção de verdadeiros positivos em relação ao número total de objetos verdadeiros. Ele indica a capacidade do modelo em encontrar todos os objetos presentes.
- A curva de precisão-*recall* é construída ao variar o limiar de confiança do modelo para determinar quais detecções são consideradas verdadeiras ou falsas. O AP é calculado ao obter a média das precisões em diferentes pontos de *recall*. Um AP maior indica um melhor desempenho na detecção de objetos.

4.3.2 Mean Average Precision (mAP):

Esta é uma média das pontuações de AP para todas as categorias de objetos. Ele fornece uma medida geral da precisão do modelo em detectar diferentes tipos de objetos. O mAP é frequentemente usado como uma métrica de referência para comparar o desempenho entre diferentes modelos ou configurações. Assim como o AP, o mAP varia de 0 a 1, sendo 1 o valor ideal.

4.3.3 Intersection over Union (IoU):

O Intersection over Union (IoU), também chamado de *Jaccard Index*, mede a sobreposição entre a caixa delimitadora prevista pelo modelo e a caixa delimitadora verdadeira do objeto. É calculado dividindo a área da interseção entre as duas caixas pela área da união das caixas.

O IoU é usado para determinar se uma detecção é considerada verdadeira positiva ou falsa positiva. Um valor de IoU acima de um limite definido é considerado uma detecção correta. Geralmente, um valor de IoU de 0,5 é usado como limiar padrão. Valores de IoU mais altos exigem uma sobreposição maior entre as caixas delimitadoras para considerar a detecção como correta.

O IoU é uma métrica importante em problemas de detecção de objetos, pois permite avaliar a precisão espacial da detecção em relação à verdadeira localização dos objetos.

Essas métricas são amplamente utilizadas para avaliar e comparar o desempenho de modelos de detecção de objetos, como o YOLO, permitindo uma análise quantitativa de sua precisão e capacidade de detecção.

5 AVALIAÇÃO E RESULTADOS

Nesse estudo, um *dataset* foi desenvolvido contendo 2.363 imagens. Destas, 2.223 foram utilizadas para treinamento do modelo preditivo e 140 para validação e aferição do desempenho do mesmo.

As imagens foram coletadas a partir da Internet em *datasets* públicos. Estas foram divididas em duas classes sendo pessoas fazendo uso da máscara corretamente e pessoas sem máscara ou com máscara colocada de forma incorreta. Decidiu-se manter apenas duas categorias em razão de facilitar a aplicação do modelo preditivo visto que o não uso de máscara ou o uso incorreto geram risco de transmissão viral.

Estes dados foram rotulados e submetidos a uma rotina de treinamento de 4.000 épocas (quantidade de ciclos de treinamento recomendados para essa quantidade de classes). O treinamento do modelo foi realizado pela plataforma GOOGLE COLAB. Ao final do treinamento foram emitidos relatórios do desempenho da rede neural. Estes relatórios são detalhados a seguir. O resultado final do modelo de treinamento está nomeado nas tabelas a seguir como *best weights*.

Tabela 1 – Matrix de confusão obtida atraves da avaliação modelo *best.weights*.

	Com Máscara (Sim)	Sem Máscara (Não)
Com Máscara (Sim)	90 (VP)	3 (FP)
Sem Máscara (Não)	2 (FN)	60 (VN)

A Tabela 1 demonstra a matriz de confusão obtida a partir dos resultados finais do modelo, em que este previu de forma correta 140 itens ante 5 casos de falsos positivos e negativos.

Na Tabela 2 é possível visualizar a evolução da *precision* e do *recall*. No fim do treinamento, o modelo obteve uma precisão de 97% e um *recall* de 99%. Uma alta assertividade demonstrada pela rede neural

Tabela 2 – Tabela demonstrativa da evolução da *precision* e *recall* a cada 1.000 epocas.

Épocas de treinamento	***Precision* VP/(VP+FP)**	***Recall* VP/(VP+FN)**
1.000	0,71	0,64
2.000	0,95	1,00
3.000	0,96	1,00
4.000	0,95	0,98
best.weights	0,97	0,99

Tabela 3 – Tabela demonstrativa da evolução do mAP a cada 1.000 épocas.

Épocas de treinamento	**AP Máscara**	**AP Sem Máscara**	**mAP**
1.000	72.14%	68.43%	70.28%
2.000	99.77%	99.89%	99.83%
3.000	99.63%	100.00%	99.82%
4.000	97.29%	100.00%	98.65%
best.weights	98.72%	100.00%	99.36%

A Tabela 3, por sua vez, apresenta a progressão da acurácia do modelo de forma individual para indivíduos com e sem máscara. Esta alcança 98,72% de acerto em imagens com máscara e 100% nas imagens sem máscara. Estes resultados demonstram o avanço da mAP que, no final do treinamento, registra 99,36%.

Tabela 4 – Tabela demonstrativa da evolução do IoU a cada 1.000 epocas.

Épocas de treinamento	**IoU**
1.000	46.16%
2.000	73.04%
3.000	77.71%
4.000	78.60%
best.weights	78.62%

A Tabela 4 mostra a evolução do IoU. No estado final do treinamento, o modelo atinge 78,62% de assertividade, refletindo na localização correta dos objetos nas imagens submetidas.

6 CONCLUSÃO

Este artigo apresentou a aplicação bem-sucedida do algoritmo YOLOv4 para o reconhecimento de pessoas com e sem máscara facial. Os resultados obtidos demonstram a eficácia do sistema, com uma precisão de 97%, *recall* de 99%, mAP de 99,36% e IoU de 78,62%.

A detecção e classificação precisas, usando YOLOv4, de pessoas mostraram-se confiáveis na distinção entre indivíduos com máscara e sem máscara. Essa capacidade é de extrema importância em cenários como a pandemia de COVID-19, em que o uso de máscaras faciais tornou-se uma medida fundamental para a manutenção da saúde pública.

A precisão de 97% do sistema significa que, na maioria dos casos, o algoritmo foi capaz de corretamente identificar se uma pessoa estava usando ou não uma máscara facial. Um *recall*

de 99% destaca a capacidade do modelo em encontrar e detectar a grande maioria das pessoas que estão sem máscara ou com máscara.

Uma mAP de 99,36% indica que o sistema alcançou uma pontuação excepcional na avaliação da precisão do modelo em todas as classes, incluindo pessoas com e sem máscara. Esse resultado demonstra que o YOLOv4 é altamente confiável na tarefa de reconhecimento de pessoas.

Além disso, o IoU de 78,62% indica uma boa sobreposição entre as caixas delimitadoras geradas pelo algoritmo e as máscaras reais utilizadas pelas pessoas. Isso significa que o sistema foi capaz de realizar uma detecção precisa e de alta qualidade, com poucos casos de falsos positivos ou falsos negativos.

Em resumo, os resultados obtidos com a aplicação do YOLOv4 para o reconhecimento de pessoas com e sem máscara mostraram-se promissores. Os números de precisão, *recall*, mAP e IoU demonstraram a capacidade do modelo em distinguir corretamente entre as duas classes. Esses resultados têm implicações significativas para a segurança e a saúde pública, especialmente em tempos de pandemia, em que o controle do uso de máscaras é crucial para prevenir a propagação de doenças.

REFERÊNCIAS

FERNANDES, A. M. **Inteligência Artificial - Noções Gerais**. Pará de Minas, MG: Virtual Books, 2004.

FIGUEIREDO, E.; SILVA, E. Combate ao covid19: Detecção em tempo real de indivíduos sem máscara em ambiente escolar por meio de deep learning. In: **XV Brazilian e-Science Workshop (BreSci)**. Porto Alegre, RS: SBC, 2021. p. 113–120.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge, MA, USA: MIT press, 2016.

HAN MICHELINE KAMBER, J. P. J. **Data Mining: Concepts and Techniques**. New York, NY, USA: Macmillan Publ., 2011.

JIANG, P. et al. A review of yolo algorithm developments. **Procedia Computer Science**, Elsevier, v. 199, p. 1066–1073, 2022.

KÄHLER, C. J.; HAIN, R. Fundamental protective mechanisms of face masks against droplet infections. **Journal of aerosol science**, Elsevier, v. 148, p. 105617, 2020.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Advances in neural information processing systems. In: ____. Cambridge, MA, USA: MIT Press, 2012. cap. ImageNet Classification with Deep Convolutional Neural Networks.

LIU, R.; REN, Z. Application of Yolo on mask detection task. In: **13th International Conference on Computer Research and Development (ICCRD)**. Virtual Conference: IEEE, 2021. p. 130–136.

MOHAPATRA, R. K. et al. The recent challenges of highly contagious COVID-19, causing respiratory infections: Symptoms, diagnosis, transmission, possible vaccines, animal models,

and immunotherapy. **Chemical Biology & Drug Design**, Wiley Online Library, v. 96, n. 5, p. 1187–1208, 2020.

NILSSON, N. J. **Introduction to Machine Learning**. 1996. <https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Introduction%20to%20Machine%20Learning%20-%20Nils%20J%20Nilsson.pdf>. Online; accessed 19 July 2023.

WANG, M.; DENG, W. Deep face recognition: A survey. **Neurocomputing**, Elsevier, v. 429, p. 215–244, 2021.

WHO. **Infection prevention and control during health care when novel coronavirus (nCoV) infection is suspected: interim guidance**. 2020. <https://covid19-evidence.paho.org/handle/20.500.12663/839>. Online; accessed 19 July 2023.

YADAV, S. Deep learning based safe social distancing and face mask detection in public areas for COVID-19 safety guidelines adherence. **International Journal for Research in Applied Science and Engineering Technology**, v. 8, p. 1368–1375, 07 2020.