

Sistema basado en inteligencia artificial para la identificación de cubrebocas y su correcto uso

Julio Cesar Elizalde-Silva, Carlos Avilés-Cruz,
Arturo Zúñiga-López

Universidad Autónoma Metropolitana,
Departamento de Electrónica, División de Ciencias Básicas e Ingeniería,
México

cesaruaeh@gmail.com,
{caviles, azl}@azc.uam.mx

Resumen. Derivado de la pandemia que se vive en México y en el mundo por el SARS-CoV-2 causante de la COVID-19, la Organización Mundial de la Salud (OMS) recomendó ampliamente el uso del cubreboca. La OMS puntualiza en el correcto uso del cubreboca para personas que se encuentran en lugares públicos, abiertos o que no puedan mantener de mínimo dos metros de distancia entre ellas. Por ende, surge la necesidad de desarrollar una herramienta computacional que permita identificar a personas en zonas públicas sin cubreboca o que no lo usen de manera correcta. El modelo que se ha propuesto en el presente trabajo utiliza técnicas de aprendizaje profundo tales como las redes neuronales convolucionales (RNC) así, como también algunos módulos de las siguientes herramientas de programación en Python: OpenCV, Scikit-learn, TensorFlow, Keras, etc. El modelo toma como entrada el video que se está transmitiendo en vivo desde una cámara, e identifica rostros de personas que no llevan puesto el cubreboca. Finalmente se muestra en la pantalla de la computadora una alarma visual, indicando que la persona no trae cubreboca o no lo usa de manera correcta. El sistema propuesto tiene una exactitud de 97 % de buen reconocimiento de cubrebocas.

Palabras clave: Covid-19, redes neuronales convolucionales, python, OpenCV, scikit-learn, tensorflow, keras, extracción de características, clasificación.

Artificial Intelligence Based System for the Identification of Mask and their Correct Use

Abstract. Derived from the pandemic that is experienced in Mexico and in the world by the SARS-CoV-2 that causes COVID-19, the World Health Organization (WHO) widely recommended the use of the mask. The WHO points out the correct use of the mask for people who are in public, open places or who cannot keep at least two meters of distance between them. Therefore, the need arises to develop a computational system that allows identifying people in public areas without a mask or who do not use it correctly. The model that has

been proposed in this work uses deep learning techniques such as convolutional neural networks (CNN) as well as some modules of the following Python programming tools: OpenCV, Scikit-learn, TensorFlow, Keras, etc. This work makes it possible to identify people with and without face masks, using CNN; consists of two main stages: a) Characteristics extraction; are the elements that describe an object (size, shape, color, texture, etc.) contain the greatest amount of information, b) Classification; At this stage is where the model learns from the training set, allowing it to distinguish between different images and thus be able to classify them from a visual representation of the same. This project can be implemented for security purposes in any hospital, clinic, school, company (public or private), etc., since its implementation is easy and efficient in terms of the resources it needs.

Keywords: Classification, content—based image retrieval, image retrieval, image classification, Wiener-Granger causality, convolutional neural network, OpenCV, scikit-learn, tensorflow, keras.

1. Introducción

Las enfermedades virales son sumamente comunes, de hecho, son las principales razones por las cuales visitamos al médico. Según la OMS, las enfermedades respiratorias son aquellas que afectan directamente las vías nasales, los bronquios y los pulmones. La mayoría de las infecciones que causan los virus se pueden prevenir si se toma en cuenta una serie de recomendaciones pertinentes; lavarse las manos con suficiente agua y jabón, evitar el contacto directo con personas infectadas y el uso correcto del cubreboca, pueden ser algunas de estas medidas.

En diciembre del 2019 hubo un brote epidémico de neumonía de causa desconocida en Wuhan, China; las autoridades chinas confirmaron 41 casos detectados entre el 8 de diciembre de 2019 y el 2 de enero de 2020. La rápida expansión de una enfermedad desconocida, hasta ese momento, hizo que la OMS la declarara emergencia sanitaria de preocupación internacional. El 11 de marzo de 2020 la enfermedad se hallaba ya en más de 100 territorios a nivel mundial [15].

Para prevenir la expansión de dicha enfermedad los gobiernos han propuesto una serie de restricciones tales como: cancelación de viajes, cancelación de eventos, cierre de establecimientos, etc. La reciente enfermedad ha tenido un alto impacto en muchos países alrededor de todo el mundo; ha cambiado de manera impresionante la vida de las personas modificando las rutinas a las cuales estaban acostumbradas. Los profesionales de la salud, hospitales, clínicas, organizaciones sanitarias e investigadores han hecho un enorme esfuerzo para sacar una vacuna que pueda ayudar a superar esta grave enfermedad.

A pesar de tener resultados exitosos, esto no es suficiente puesto que el virus se propaga a través del aire. Cuando una persona infectada habla o estornuda las gotas que salen de su boca o nariz se diseminan por el aire y así, afecta a otras personas [14], por lo que usar cubreboca puede ser una medida de gran importancia para disminuir el riesgo de ser contagiado.

La inteligencia artificial (IA) comprende un amplio conjunto de algoritmos, y su efectividad está estrechamente relacionada con la calidad de la información de la cual se aprende [16]. El aprendizaje profundo es uno de los campos de la IA, que más ha despertado el interés en científicos e investigadores debido a su enorme campo de aplicación [17]. La detección de objetos en imágenes digitales es quizá la tarea más importante del aprendizaje profundo y, la técnica más utilizada para llevar a cabo dicha tarea son las RNC que han logrado resultados prometedores.

Con el enorme desarrollo que han tenido estas técnicas, la detección de rostros en imágenes digitales, parece ser un problema que se ha solventado de manera correcta [3], sin embargo, la detección de rostros humanos con objetos tales como: lentes, pasamontañas, gorras, sombreros, cubreboca etc., hacen que los detectores de rostros sean cada vez más complejos y que su implementación, no proporcione los resultados esperados.

En el presente trabajo hemos propuesto un sistema que identifica a personas que no portan cubreboca o que no lo usan de manera correcta. El sistema hace uso de un modelo de aprendizaje profundo, el cual está compuesto de una RNC de dos capas internas, una capa que recibe todos los atributos obtenidos de las capas anteriores, una capa interna que evita el sobre entrenamiento del modelo, una capa oculta compuesta por 50 valores de salida y finalmente una capa de salida con dos clases.

El trabajo propuesto en este artículo también hace uso de un método de la librería OpenCV que se llama “detectMultiScale” de la interfaz “CascadeClassifier” la cual nos permite identificar el rostro de una o más personas de manera frontal. Este método permite obtener un rectángulo delimitador, donde se encuentra la cara de la (s) persona (s) en la imagen. Para entrenar el modelo se utilizaron 1376 imágenes de rostros humanos; de las cuales 690 imágenes son de personas que traen puesto el cubreboca y 686 imágenes son de personas que no traen cubreboca.

Para reducir el costo computacional que implica el preprocesamiento de la información, las imágenes fueron convertidas a escala de grises, así, como también se redujo el tamaño de estas quedando de 32×32 . Finalmente, las imágenes en su representación matricial fueron normalizadas para comprimir aún más los datos que entran al modelo. Los datos obtenidos han sido divididos con el método “train_test_split” de la librería Scikit-learn, utilizando el 80 % para entrenamiento, 20 % para prueba y dejando la configuración del método de manera aleatoria. La exactitud del modelo es de 97 % con 20 épocas.

El resto del trabajo está organizado de la siguiente manera: en la sección 2, se describen los trabajos que se han llevado a cabo con respecto a esta área de investigación, en la sección 3, se explica la metodología propuesta que se llevó a cabo, en la sección 4, se muestran los resultados obtenidos. Finalmente, en la sección 5, se ponen las conclusiones y el trabajo futuro.

2. Trabajos relacionados

Los algoritmos convencionales de detección de objetos fueron principalmente derivados del aprendizaje automático; siendo este una disciplina que construye modelos matemáticos y algoritmos para realizar tareas específicas utilizando computadoras [12].

Los científicos de la computación han desarrollado numerosas arquitecturas que ayudan a identificar y clasificar objetos en una imagen digital, utilizando técnicas de aprendizaje profundo. El aprendizaje profundo ha mostrado un enorme potencial en muchas aplicaciones en la vida científica y práctica, la detección de rostros, texto, logotipos, video, vehículos, imágenes médicas; son algunas de estas aplicaciones [11].

Proporciona un grupo de algoritmos que se pueden emplear para problemas de aprendizaje supervisado, no supervisado y/o reforzado, sobre cualquier tipo de datos, señales, imágenes digitales, videos, etc.

La idea fundamental del aprendizaje profundo proviene de neuronas artificiales, el cual es un modelo de aprendizaje jerárquico muy robusto que permite aprender representaciones complejas directamente de los datos de entrada, el aprendizaje profundo ha ido de la mano de las RNC obteniendo resultados prometedores.

En esencia la identificación y clasificación de objetos es utilizar cuadros rectangulares delimitadores para localizar los objetos en la imagen [20]. Se compone principalmente de cuatro etapas:

1. Identificación de objetos; obtiene la categoría a la que pertenecen los objetos en la imagen.
2. Detección de objetos; ubica los objetos con cuadros delimitadores rectangulares.
3. Segmentación semántica; predice las categorías de cada píxel.
4. Segmentación de instancias; necesita predecir tanto las categorías de cada píxel como de cada objeto [13].

Las RNC han tenido una evolución significativa en la clasificación de imágenes[19] siendo el tipo de aprendizaje supervisado preferido para problemas de visión por computadora. Sin embargo, diseñar mejores arquitecturas de RNC sigue siendo una pregunta inicial. A continuación, se muestra una breve descripción de los algoritmos más utilizados de RNC.

- Detección de objetos: Probablemente es la tarea más importante de la visión por computadora. Esta técnica utiliza una ventana de tamaño $M \times N$ que selecciona el objeto en la imagen, los objetos ubicados son encerrados en cuadros delimitadores [7]. Inicialmente, un clasificador se prepara con un conjunto de datos de entrenamiento (imágenes), si encuentra el objeto de interés, lo marca como una imagen positiva, de lo contrario como una imagen negativa.
- Faster R-CNN: Comprende dos módulos, la red de propuesta de región (RPN), en la que se distingue la región del objeto en la imagen y una red que permite clasificar los objetos en la región propuesta [1]. Faster R-CNN es considerada una de las técnicas de detección de objetos más precisas.
- YOLO – You Only Look Once: Una sola RNC predice simultáneamente múltiples cuadros delimitadores y probabilidades de clase para detectar objetos en imágenes [18]. Esta técnica divide la imagen en regiones, poniendo cuadros de identificación y probabilidades por cada región. Los cuadros son ponderados a partir de las probabilidades predichas y utiliza sus funciones para predecir cada cuadro delimitador [6].

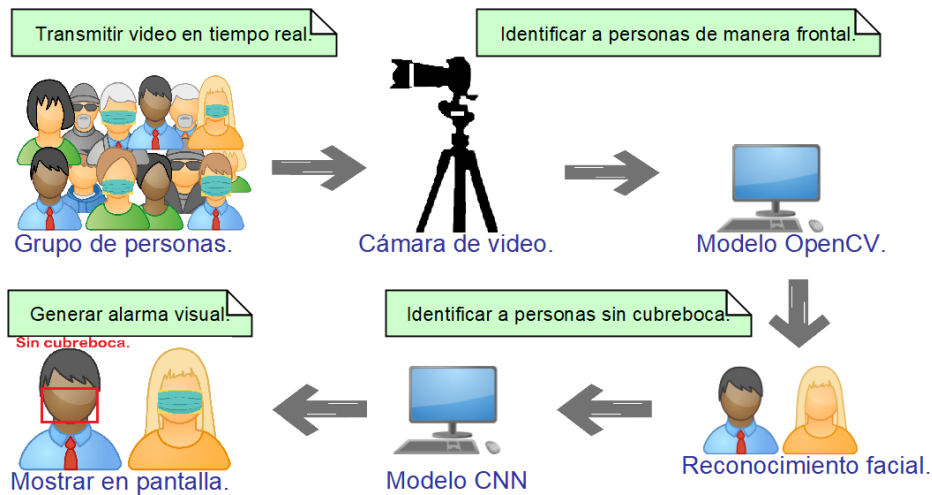


Fig. 1. Metodología general de sistema de identificación de cubreboca.

- Single shot detector (SSD): Encuentra numerosos objetos en la imagen con tan solo una muestra. La RNC se ejecuta una sola vez obteniendo el mapa de características de la imagen. Utiliza rectángulos delimitadores para predecir objetos que tienen diferente tamaño, la RNC combina n predicciones de mapas de características con diferentes resoluciones. El núcleo de SSD es predecir puntuaciones de cada categoría para un conjunto fijo de cuadros delimitadores predeterminados que utilizan pequeños filtros convolucionales aplicados a los mapas de características [10].
- RetinaNet: Es una técnica que está compuesta de dos arquitecturas: YOLO y SSD, cuenta con redes piramidales para la detección de objetos. RetinaNet es un modelo unificado, compuesto por una red que utiliza como columna vertebral y dos subredes específicas de tareas. La red que utiliza como columna vertebral se encarga de obtener las características de la imagen de entrada. La primera subred realiza la clasificación de objetos. La segunda subred crea un cuadro rectangular delimitador, utilizando la salida de la red troncal [5].

Como se ha mencionado anteriormente las RNC han demostrado gran capacidad para resolver problemas de clasificación de imágenes, este aprendizaje se obtiene gracias a la enorme cantidad de información (millones de parámetros) que utilizan las redes para su entrenamiento. Por otro lado, el reconocimiento facial representa un enorme campo de aplicación para estos modelos. La detección de rostros se refiere al uso de un método para determinar la posición del rostro en cualquier imagen o video [21].

Los detectores de rostros modernos pueden identificar fácilmente el rostro de una persona de manera frontal[9]. Sin embargo, la precisión de la detección de rostros puede verse afectada por varios factores ajenos al modelo, ejemplo: iluminación desigual, elementos extraños, los rostros están de perfil, ángulos complicados, oclusión, etc. La investigación sobre detección y reconocimiento facial se ha estudiado ampliamente en los últimos años.

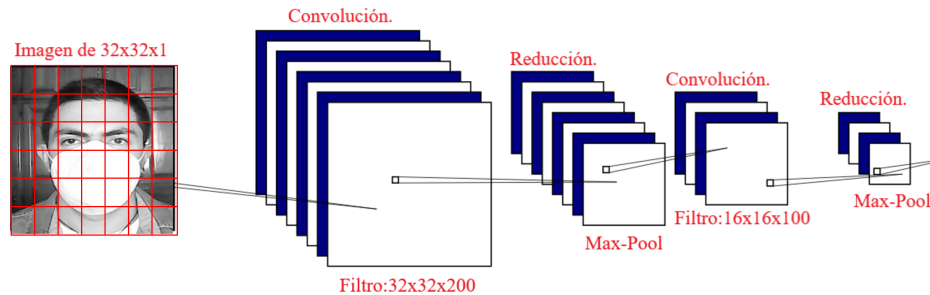


Fig. 2. Red neuronal convolucional.

Las aplicaciones de reconocimiento facial juegan un papel importante en muchas áreas como la seguridad, vigilancia con cámaras, verificación de identidad en dispositivos electrónicos modernos, investigaciones criminales, sistemas de gestión de bases de datos, aplicaciones de tarjetas inteligentes, etc. [2] Algunas de las ventajas que tienen los sistemas de reconocimiento facial son: utilizan medios no invasivos, pueden encontrar a una persona dentro de una gran base de datos, identificar a personas concretas en tiempo real, etc. [8].

El objetivo principal del reconocimiento facial es autenticar e identificar los rasgos faciales de una persona. Sin embargo, los rasgos faciales se capturan en tiempo real y son procesados por un modelo previamente entrenado, las RNC han demostrado ser un enfoque altamente eficiente para esta tarea[4]; para lograr esta precisión óptima el conjunto de datos de entrenamiento es elevado.

3. Sistema propuesto

La metodología general del sistema propuesto se muestra en la figura 1. La propuesta inicia con: a) La captura de video en tiempo real a través de una cámara óptica, b) Identificación de rostros de manera frontal, c) Red neuronal convolucional, y d) Mostrar resultados en pantalla. Los módulos son descritos a continuación:

Etapa de entrenamiento

– Aumento de datos:

El proyecto está compuesto por el módulo “DataAugmentation”, encargado de aumentar los datos que son utilizados para el entrenamiento del modelo, las imágenes son leídas de las carpetas “mask” y “without_mask”. La clase se llama “ImageDataGenerator” e implementa el método “flow”, su tarea principal consiste en cambiar las propiedades de cada imagen obtenida, generar una nueva imagen y finalmente guardar las imágenes generadas en las mismas carpetas de donde se obtuvieron.

– Preprocesamiento de la información:

Viene la etapa de preprocesamiento de la información, es decir las imágenes que fueron generadas ahora van a hacer los insumos de nuestro modelo. La clase encargada de llevar a cabo esta tarea se llama “DataIn” y su tarea inicial es leer

las imágenes de los directorios “mask” y “without_mask”, convertirlas a escala de grises y así poder obtener un solo canal que va desde 0 hasta 255. Después de haber realizado este primer paso, también se ha disminuido el tamaño de cada imagen, quedando de la siguiente manera: $32 \times 32 \times 1$, lo cual equivale a tener una matriz de 1024 valores (píxeles) o lo que es lo mismo tener una matriz de 32 filas por 32 columnas. Los datos han sido normalizados entre 0 y 1 para trabajar con valores más pequeños. Finalmente usando el método “to_categorical” de Keras obtenemos una matriz binaria etiquetando con 0 a las personas que traen puesto el cubreboca y con 1 a personas sin cubreboca.

– **Imágenes de entrenamiento:**

El módulo central del proyecto se llama “Classify” y su tarea inicial consiste en obtener los datos de entrenamiento y prueba que ocupa el modelo. Como se mencionó anteriormente los datos están divididos en dos partes: 80 % para entrenamiento y 20 % para prueba; esta tarea se llevó a cabo haciendo uso del método “train_test_split” del api de Scikit-learn y dejando la configuración del método de manera aleatoria. Para la extracción de características se ha utilizado una RNC secuencial con dos capas internas. El diagrama se puede ver en la figura [2].

– **Parámetros de entrenamiento de la RNC:**

Como se muestra en el diagrama la RNC toma como entrada una imagen de 32×32 de un solo canal (escala de grises), para el proceso de convolución se han utilizado 200 filtros de 3×3 y una función de activación ReLu, finalmente para obtener la información más representativa se configuro un submuestreo al máximo de 2×2 (Max-Pool) y con esto se da por terminada la configuración de la capa 1.

Para el proceso de convolución de la capa 2, se tomaron 100 filtros de 3×3 , un “padding same” para no reducir el tamaño de la imagen y una función de activación ReLu. Para obtener la información más representativa de las imágenes se utilizó un Max-Pool de 2×2 , así, como un stride de 2×2 que minimiza aun más el tamaño de las imágenes.

– **Etapa de clasificación de la RNC:**

Para la etapa de clasificación, se cuenta con una capa completamente conectada, por sus siglas en ingles “Full-connected layer” que recibe todos los atributos extraídos de las capas anteriores y se colocan en un vector $N \times 1$, donde N es igual al número de atributos; en nuestro caso igual a 6,400. Para evitar el sobre entrenamiento del modelo (overfitting), utilizamos una capa de regularización (Dropout), inhabilitando el 20 % de los parámetros.

El modelo cuenta con dos capas internas de salida; la primera está configurada para tomar 50 valores o lo que es lo mismo tener 50 clases diferentes, además de una función de activación ReLu; en la segunda capa el número de valores ha sido reducido, quedando el modelo con 2 clases, se utilizó una función de activación “Softmax”, con la cual forzamos a tener una sola clase ganadora, es decir la clase que tenga mayor probabilidad entre 0 y 1 es la clase que gana. En la figura [3] se muestra el diagrama de la capa “Full-connected layer”.

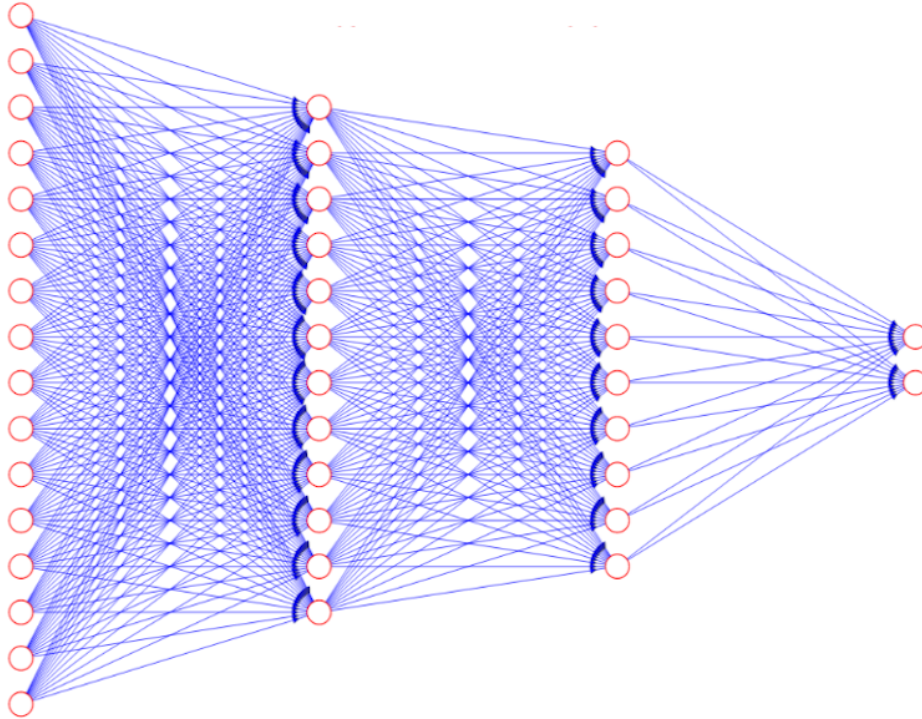


Fig. 3. Capa completamente conectada.

Para optimizar el costo del modelo se utilizó la función del descenso del gradiente estocástico “adam”, dejando su configuración por default, así como la medida de precisión “entropía cruzada binaria”, utilizada para ajustar los pesos del modelo durante el entrenamiento. Finalmente, el modelo ha sido entrenado, obteniendo el 97 % de exactitud con 20 épocas.

El ultimo módulo del sistema se llama “DataOut”, su tarea inicial consiste en cargar en el disco de la computadora el modelo obtenido en el paso anterior, posteriormente se utiliza la interfaz “VideoCapture” del api de OpenCV, la cual toma el control de la cámara y la entrada del modelo es el resultado de la interfaz “VideoCapture”, la cual consta de un video en escala de grises que se transmite en vivo desde la cámara interna de la computadora.

– **Detección de rostros en la imagen:**

Para detectar el rostro de una persona, se utiliza el método “detectMultiScale” de la interfaz “CascadeClassifier”, utilizando un rectángulo delimitador de color verde donde se encuentra el rostro de la (s) persona (s) identificadas en el video. Finalmente se muestra en la pantalla de la computadora una alarma visual, indicando que la persona no trae puesto el cubreboca. Para finalizar, el sistema libera los recursos de la computadora, es decir, destruye la ventana emergente que aparece en la pantalla y libera la cámara de la computadora.

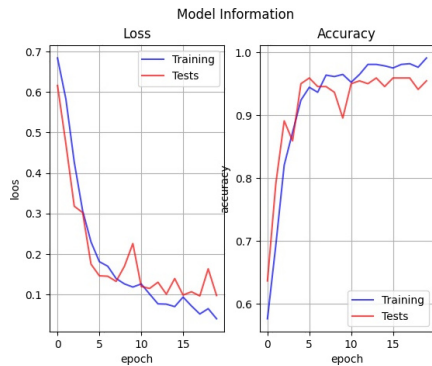


Fig. 4. Perdida y exactitud.



Fig. 5. Personas capturadas en tiempo real.

4. Resultados

Para probar la metodología propuesta, así como evaluar su eficiencia, se utilizó la base de imágenes del repositorio¹. La base de imágenes esta conformada de 1,376 imágenes, de las cuales 690 son de rostros de personas con cubreboca y 686 son de rostros de personas sin cubreboca. El trabajo se realizó sobre el lenguaje de programación Python, utilizando además las principales API's de aprendizaje automático (TensorFlow, Keras y Scikit-learn) que ayudan a la implementación del modelo propuesto, así como también OpenCV que permite la manipulación de imágenes e implementación de la interfaz "CascadeClassifier" para detectar el rostro de las personas de manera frontal.

El modelo propuesto en el presente trabajo, fue entrenado con 20 épocas. La figura 4 muestra las gráficas de pérdida y exactitud respectivamente. Como se puede apreciar en dicha gráfica, el error (gráfica izquierda) de entrenamiento es menor al 0,5 % (ver trazo azul). Por otra parte, la precisión alcanza un desempeño del 98 % (ver trazo azul, de la gráfica izquierda). En cuanto al desempeño global, usando toda la base de imágenes, se logra una exactitud del 97 %. En la figura 5 se muestra un ejemplo de la detección correcta del cubreboca, así como la detección de la persona que no porta el cubreboca. Como se puede apreciar en la figura 5, la persona del lado derecho no porta el cubreboca, indicándolo en un recuadro rojo y con el texto respectivo.

5. Conclusiones y perspectivas

El sistema que se ha desarrollado en el presente trabajo es una medida de seguridad de las muchas que recomienda la OMS para disminuir el riesgo de contagio de la COVID-19, sirve para identificar a personas en tiempo real que no traen cubreboca o que no lo usan de manera correcta, esta herramienta podría permitir automatizar el proceso que se lleva a cabo en establecimientos públicos, donde hay una persona encargada de revisar que las personas que quieren ingresar al lugar lleven puesto su cubreboca o que lo usen de manera correcta.

¹ <https://github.com/prajnasb/observations/tree/master/experiements/data>

El sistema propuesto proporciona una eficiencia del 97 %, presentando un recuadro rojo en pantalla con el mensaje “sin cubreboca” cuando la persona no lleva puesto su cubreboca o no lo use de manera correcta. Una ventaja del sistema es que trabaja en tiempo real (procesar al menos 24 imágenes por segundo) en una computadora estándar.

Una limitante del sistema propuesto es que solo sirve para detectar a personas de manera frontal. Para trabajo futuro se ha pensado en complementar el sistema; el cual permitirá identificar el rostro de una persona, no solo de manera frontal, sino también de perfil, además de detectar la temperatura corporal utilizando una cámara termográfica.

Referencias

1. Ajitha, S., Judy, M. V.: Faster R-CNN classification for the recognition of glaucoma. Paper presented at the Journal of Physics: Conference Series, vol. 1706, no. 1, pp. 1–12 (2020) doi: 10.1088/1742-6596/1706/1/012170
2. Asif Hussain, S., Al Balushi, A. S. A.: A real time face emotion classification and recognition using deep learning model. Journal of Physics: Conference Series, vol. 1432, no. 1, pp. 14 (2020) doi: 10.1088/1742-6596/1432/1/012087
3. Basha, C. Z., Pravallika, B. N. L., Shankar, E. B.: An efficient face mask detector with pytorch and deep learning. European Alliance for Innovation, Endorsed Transactions on Pervasive Health and Technology, vol. 7, no. 25, pp. 1–8 (2021) doi: 10.4108/eai.8-1-2021.167843
4. Ben-Fredj, H., Bouguezzi, S., Souani, C.: Face recognition in unconstrained environment with cnn. Visual Computer, vol. 37, pp. 217–226 (2021) doi: 10.1007/s00371-020-01794-9
5. Dhillon, A., Verma, G. K.: Convolutional neural network: A review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, vol. 9, no. 2, pp. 85–112 (2020) doi: 10.1007/s13748-019-00203-0
6. Inthiyaz, S., Ahammad, S. H., Sai Krishna, A., Bhargavi, V., Govardhan, D., Rajesh, V.: YOLO (you only look once) making object detection work in medical imaging on convolution detection system. International Journal of Pharmaceutical Research, vol. 12, no. 2, pp. 312–326 (2020) doi: 10.31838/ijpr/2020.12.02.0003
7. Kang, K., Ouyang, W., Li, H., Wang, X.: Object detection from video tubelets with convolutional neural networks. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 817–825 (2016) doi: 10.1109/CVPR.2016.95
8. Lawrence, S., Giles, C. L., Tsoi, A. C., Back, A. D.: Face recognition: A convolutional neural-network approach. IEEE Transactions on Neural Networks, vol. 8, no. 1, pp. 98–113 (1997) doi: 10.1109/72.554195
9. Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G.: A convolutional neural network cascade for face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5325–5334 (2015) doi: 10.1109/CVPR.2015.7299170
10. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C.: SSD: Single shot multibox detector. Lecture Notes in Computer Science, pp. 21–37 (2016) doi: 10.1007/978-3-319-46448-0_2
11. Loey, M., Manogaran, G., Taha Mohamed, H. N., Nour Eldeen, M. K.: Fighting against COVID-19: A novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection. Sustainable Cities and Society, vol. 65, pp. 1–8 (2021) doi: 10.1016/j.scs.2020.102600
12. Masita, K. L., Hasan, A. N., Shongwe, T.: Deep learning in object detection: A review. In: Proceedings of the International Conference on Artificial Intelligence, Big Data, Computing

- and Data Communication Systems (icABCD), pp. 11 (2020) doi: 10.1109/icABCD49160.2020.9183866
13. Meenpal, T., Balakrishnan, A., Verma, A.: Facial mask detection using semantic segmentation. In: *Proceedings of the 4th International Conference on Computing, Communications and Security*, pp. 1–5 (2019) doi: 10.1109/CCCS.2019.8888092
 14. Megahed, N. A., Ghoneim, E. M.: Antivirus-built environment: Lessons learned from COVID-19 pandemic. *Sustainable Cities and Society*, vol. 61, pp. 23 (2020) doi: 10.1016/j.scs.2020.102350
 15. Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., Hemanth, J.: SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustainable Cities and Society*, vol. 66, pp. 11 (2021) doi: 10.1016/j.scs.2020.102692
 16. Peter Norvig, S. J. R.: *Artificial Intelligence: A Modern Approach* (1995)
 17. Ponce-Gallegos, J. C., Torres-Soto, A., Quezada-Aguilera, F. S., Silva-Sprock, A., Martínez-Flor, E. U., Casali, A., Scheihing, E., Túpac-Valdivia, Y. J., Torres-Soto, M. D., Ornelas-Zapata, F. J., Hernández, J. A., Zavala, C., Vakhnia, N., Pedreño, O.: *Inteligencia Artificial. Iniciativa Latinoamericana de Libros de Texto Abiertos (LATIn)* (2014) doi: 10.13140/2.1.3720.0960
 18. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788 (2016), doi: 10.1109/CVPR.2016.91
 19. Saca, F., Ramírez, A., Cruz, C., Villegas Cortez, J.: Red neuronal convolucional con extracción de características multi-columna para clasificación de imágenes. *Research in Computing Science*, vol. 148, no. 7, pp. 391–404 (2019) doi: 10.13053/rcs-148-7-29
 20. Xiao, Y., Tian, Z., Yu, J., Zhang, Y., Liu, S., Du, S., Lan, X.: A review of object detection based on deep learning. *Multimedia Tools and Applications*, vol. 79, no. 34, pp. 23729–23791 (2020) doi: 10.1007/s11042-020-08976-6
 21. Yan, H., Wang, X., Liu, Y., Zhang, Y., Li, H.: A new face detection method based on faster RCNN. *Journal of Physics: Conference Series*, vol. 1754, pp. 1–6 (2021) doi: 10.1088/1742-6596/1754/1/012209