



## **PROYECTO TERCER PARCIAL**

### **Preprocesamiento y Segmentación.**

UPIIT-IPN

21 de junio 2024

Visión Artificial

Italy Abril Zayas Riojas

Homero Meneses Vázquez

## Resumen

Presentamos un nuevo conjunto de datos desafiante, CPPE-5 (Equipo de Protección Personal Médico), con el objetivo de permitir el estudio de la categorización subordinada de equipos de protección personal médico, algo que no es posible con otros conjuntos de datos populares que se centran en categorías amplias (como PASCAL VOC, ImageNet, Microsoft COCO, OpenImages, etc.). Para facilitar el uso de los modelos entrenados en este conjunto de datos en escenarios prácticos y complejos, nuestro conjunto de datos contiene principalmente imágenes que muestran escenas complejas con varios objetos en su contexto natural. Nuestro conjunto de datos incluye 5 categorías de objetos (monos, protectores faciales, guantes, mascarillas y gafas), y cada imagen está anotada con un conjunto de cuadros delimitadores y etiquetas positivas. Presentamos un análisis detallado del conjunto de datos en comparación con otros conjuntos de datos populares de categorías amplias, así como conjuntos de datos centrados en equipos de protección personal. También analizamos el rendimiento y comparamos la complejidad de los modelos en modelos base y de última generación para los resultados de los cuadros delimitadores.

**Palabras clave:** Conjunto de datos de referencia · Visión por computadora · Detección de objetos

## Introducción

El aprendizaje profundo está revolucionando múltiples áreas de la visión por computadora. La popularidad explosiva en este campo surgió después del Desafío de Reconocimiento Visual a Gran Escala de ImageNet (ILSVRC) y ha impulsado el estado del arte en la detección genérica de objetos. Este desafío incluye una prueba de detección utilizando imágenes de ImageNet. Desde entonces, el rendimiento de los modelos ha mejorado a velocidades sin precedentes. Entre los muchos desafíos en el aprendizaje automático, la recopilación de datos se está convirtiendo en uno de los cuellos de botella críticos. A medida que el aprendizaje profundo se vuelve popular, el núcleo de su éxito radica en la necesidad de datos de entrenamiento ricos y grandes anotados. Conjuntos de datos más grandes y ricos han sido fundamentales para la investigación de vanguardia en visión por computadora, permitiendo la próxima generación de algoritmos de última generación y han sido instrumentales en el progreso de la detección de objetos en la última década.

La detección de objetos es un problema fundamental de la visión por computadora que se ocupa de detectar instancias de objetos visuales de una cierta clase en imágenes digitales. El objetivo de la detección de objetos es desarrollar modelos y técnicas para proporcionar la información: "¿qué objetos están y dónde?". Construir conjuntos de datos más grandes y ricos juega un papel clave en permitir que las computadoras identifiquen e interpreten imágenes como composiciones de uno o varios objetos, lo que hasta ahora ha sido bastante complicado para las máquinas. A través de este conjunto de datos de detección de objetos, nuestro objetivo principal es avanzar en las máquinas para que identifiquen automáticamente dónde se encuentran los objetos (equipos de protección personal) con precisión.

En la detección de objetos, se han lanzado numerosos conjuntos de datos y benchmarks conocidos en los últimos 10 años. La mayoría de los conjuntos de datos contienen una amplia variedad de clases comunes, como diferentes tipos de animales o cosas inanimadas. Varios de estos conjuntos de datos se han convertido en estándares para la comunidad, incluyendo MIT-CSAIL, PASCAL VOC Challenges (e.g., VOC2007, VOC2012), ImageNet, Caltech-256, Microsoft COCO y DOTA. Sin embargo, este conjunto de datos fue construido con la idea de permitir la categorización subordinada, especialmente para detectar equipos de protección personal, lo que no es posible con otros conjuntos de datos populares a gran escala que se centran en categorías más amplias. Aunque la primera parte del subconjunto del conjunto de datos se lanzó para facilitar el trabajo en equipos de protección personal médico, estos se portaron cuidadosamente para crear el conjunto de datos final ampliando los objetivos a equipos de protección personal médico.

El COVID-19 está causando una morbilidad y mortalidad generalizadas a nivel mundial. El síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2) responsable de esta enfermedad infectó a más de 17 millones de personas en agosto de 2020. También se ha observado que la tendencia global es aproximadamente exponencial, a una tasa de 10 veces cada 19 días. Teniendo esto en cuenta, es muy importante poder detectar con precisión el Equipo de Protección Personal Médico para ayudar a limitar el crecimiento de COVID-19. Para fomentar el desarrollo de tales herramientas, presentamos este conjunto de datos públicamente en GitHub y un subconjunto del conjunto de datos en Kaggle, enfocándonos en identificar con precisión los Equipos de Protección Personal a través de imágenes.

En este artículo, presentamos el CPPE-5 (Equipo de Protección Personal Médico), un conjunto de datos de detección de objetos, que contiene imágenes y anotaciones de referencia para la tarea de detección de objetos. La mayoría de las imágenes se han recopilado de Flickr, con el objetivo de recolectar una mayoría de imágenes no icónicas. Una pequeña porción de imágenes se recopiló también de Google Images. Después de esto, cada una de las imágenes fue anotada utilizando técnicas de crowd-sourcing. Cada una de estas anotaciones fue evaluada por múltiples personas y luego evaluada por nosotros para mantener un estricto control de calidad de las anotaciones de referencia.

Como se mencionó, proporcionamos anotaciones unificadas para la tarea de detección de objetos con el conjunto de datos. En la Figura 1 mostramos ejemplos de anotaciones proporcionadas en el conjunto de datos. En (a) monos, guantes, mascarilla, gafas (b) monos, guantes, mascarilla (c) monos, guantes, gafas (d) monos, guantes, mascarilla, gafas (e) monos, guantes, mascarilla, gafas, protector facial (f) monos, mascarilla, gafas están demostrados. Algunas imágenes de muestra adicionales para cada categoría se pueden encontrar en el Apéndice B.

Con el conjunto de datos CPPE-5, esperamos facilitar la investigación y el uso en aplicaciones en múltiples lugares públicos para identificar de manera autónoma si se ha usado un kit de PPE y también qué parte del kit de PPE se ha usado. Uno de los objetivos principales de este conjunto de datos también fue capturar una mayor proporción de imágenes no icónicas o perspectivas no canónicas de los objetos en este conjunto de datos. Esperamos ver un alto uso de este conjunto de datos para ayudar en escenarios médicos, lo que tendría un gran efecto a nivel mundial.

El resto de este artículo está organizado de la siguiente manera: en la Sección 2 se presentan los trabajos relacionados. En la Sección 3 describimos el proceso utilizado para recopilar y anotar el conjunto de datos. En la Sección 4 presentamos estadísticas relacionadas con el conjunto de datos. En la Sección 5 presentamos los resultados experimentales, entrenando múltiples modelos de última generación y modelos base. En la Sección 6 concluimos el artículo y damos trabajos futuros.

## 2. Trabajos relacionados

A lo largo de la historia de la investigación en visión por computadora, los conjuntos de datos ricos y grandes han jugado un papel muy importante. No solo proporcionan un medio para entrenar y evaluar algoritmos, sino que también impulsan la investigación en nuevas y más desafiantes direcciones. Conjuntos de datos anteriores como el Conjunto de Datos de Categoría de Objetos Caltech-256 y la Base de Datos de Peatones de MIT facilitaron la comparación directa de cientos de algoritmos de visión por computadora y también impulsaron problemas más complejos. Conjuntos de datos recientes como el conjunto de datos Open Images v4 con aproximadamente 9.2 millones de imágenes, el conjunto de datos ImageNet con aproximadamente 14 millones de imágenes y Microsoft COCO con aproximadamente 2.5 millones de instancias etiquetadas han permitido avances en la investigación de detección de objetos con una nueva ola de algoritmos de aprendizaje profundo.

Realizar detección de objetos a menudo requiere identificar a qué clase específica pertenece el objeto y también localizar el objeto en la imagen, generalmente con un cuadro delimitador como se muestra en la Figura 1. Uno de los primeros algoritmos se centró en la detección de rostros, a menudo utilizando conjuntos de datos ad hoc. Más tarde se construyeron conjuntos de datos más realistas y desafiantes que facilitaron la creación de muchos algoritmos de aprendizaje profundo. Los Transformers fueron introducidos por primera vez en visión en Vision Transformer (ViT) dividiendo una imagen en una secuencia de tokens visuales. La estrategia de autoatención en ViTs ha demostrado un rendimiento superior en comparación con las redes neuronales convolucionales modernas (ConvNets) cuando se entrenan con recetas optimizadas. Mucha popularidad en el uso de Transformers para tareas de detección de objetos se trajo a través de DETection TRansformer (DETR) y logró resultados a la par con métodos anteriores como Faster RCNN. Después de esto, múltiples trabajos intentaron entrenar transformers para detección de objetos utilizando principalmente ViTs directamente para detección de objetos.

### Código

#### Importaciones:

- `from google.colab.patches import cv2_imshow`: Importa una función específica para mostrar imágenes dentro de los cuadernos de Google Colab usando OpenCV.
- `import cv2`: Importa la biblioteca OpenCV, que proporciona un conjunto completo de funciones para el procesamiento de imágenes y tareas de visión por computadora.

- `import numpy as np`: Importa la biblioteca NumPy, una biblioteca fundamental para la computación científica en Python, especialmente para trabajar con arreglos.
- `import tensorflow as tf`: Importa la biblioteca TensorFlow, un marco poderoso para el aprendizaje automático y el aprendizaje profundo.
- `from tensorflow.keras.applications import MobileNetV2`: Importa el modelo pre-entrenado MobileNetV2, una arquitectura de red neuronal convolucional (CNN) comúnmente utilizada para la clasificación de imágenes y la extracción de características.
- `from tensorflow.keras.models import Model`: Importa la clase Model de la API Keras de TensorFlow, utilizada para crear y administrar modelos de aprendizaje profundo personalizados.
- `from tensorflow.keras.layers import Dense, GlobalAveragePooling2D`: Importa dos capas de Keras: Dense para capas completamente conectadas y GlobalAveragePooling2D para el agrupamiento promedio global.

### **Carga y preprocesamiento de imágenes:**

- `def load_and_preprocess_image(path)::` Esta función define cómo se cargan y preprocesan las imágenes.
  - `image = cv2.imread(path)`: Lee una imagen de la ruta de archivo especificada usando la función `imread` de OpenCV.
  - `image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)`: Convierte la imagen del formato BGR (formato de color predeterminado de OpenCV) al formato RGB, que es más común para los modelos de aprendizaje profundo.
  - `image = cv2.resize(image, (224, 224))`: Cambia el tamaño de la imagen a un tamaño fijo de 224x224 píxeles, coincidiendo con el tamaño de entrada esperado por el modelo MobileNetV2.
  - `image = image / 255.0`: Normaliza los valores de píxel en la imagen para que estén entre 0 y 1, lo cual es una práctica común para los modelos de aprendizaje profundo.
  - `return image`: Devuelve la imagen preprocesada.

### **Definición de la Red de propuesta de región (RPN) (No utilizada en este fragmento de código):**

- `def create_rpn(base_layers, num_anchors)`: Esta función define una RPN (Red de propuesta de región) simple, pero no se utiliza en el fragmento de código proporcionado. Las RPN se utilizan típicamente en modelos de detección de objetos para generar posibles cuadros delimitadores (regiones de interés) para objetos en una imagen.

### **Modelo de extracción de características:**

- `def feature_extraction_model()`: Esta función define un modelo para extraer características de las imágenes.
  - `base_model = MobileNetV2(weights='imagenet', include_top=False, input_shape=(224, 224, 3))`: Crea un modelo MobileNetV2 pre-entrenado en el conjunto de datos ImageNet, utilizando pesos de modelos pre-entrenados para aprovechar el

conocimiento existente para la extracción de características. El argumento `include_top=False` excluye las capas de clasificación finales de MobileNetV2, ya que solo estamos interesados en sus capacidades de extracción de características. La forma de entrada se establece en `224x224x3` para coincidir con las expectativas del modelo pre-entrenado.

- `x = base_model.output`: Toma la salida del modelo MobileNetV2 como punto de partida para capas adicionales.
- `x = GlobalAveragePooling2D()(x)`: Aplica el agrupamiento promedio global, que reduce las dimensiones espaciales (ancho y alto) de los mapas de características a un solo valor para cada canal. Esto ayuda a capturar las características generales de la imagen.
- `feature_extractor = Model(inputs=base_model.input, outputs=x)`: Crea un nuevo modelo Keras que toma la entrada del modelo MobileNetV2 (`base_model.input`) y genera las características agrupadas promedio globalmente (`x`).
- `return feature_extractor`: Devuelve el modelo de extracción de características.

### **Carga y preparación de datos:**

- `image_paths = ["/content/snap07071.jpg", "/content/21fdf.jpg"]`: Define rutas de ejemplo para dos imágenes

### **Conclusión**

El conjunto de datos CPPE-5 (Equipo de Protección Personal Médico) es un nuevo recurso valioso para la investigación en detección de objetos, particularmente en el contexto de equipos de protección personal para la salud. Este conjunto de datos presenta varias características que lo hacen único y valioso:

- **Enfatiza la categorización subordinada:** A diferencia de otros conjuntos de datos populares que se centran en categorías amplias, CPPE-5 se enfoca específicamente en la identificación de cinco tipos de equipos de protección personal (monos, protectores faciales, guantes, mascarillas y gafas) en imágenes complejas. Esto permite un análisis más granular de la detección de objetos en este dominio específico.
- **Contiene imágenes no icónicas y perspectivas no canónicas:** CPPE-5 se esfuerza por capturar una mayor proporción de imágenes no icónicas o perspectivas no canónicas de los objetos, lo que lo hace más realista y desafiante que otros conjuntos de datos. Esto permite entrenar modelos más robustos que puedan generalizarse mejor a escenarios del mundo real.
- **Anotaciones de alta calidad:** Las imágenes en CPPE-5 están cuidadosamente anotadas con cuadros delimitadores y etiquetas positivas para cada objeto de interés. Estas anotaciones fueron revisadas por múltiples personas para garantizar la precisión y la calidad.
- **Facilita la investigación y aplicaciones:** El conjunto de datos está disponible públicamente en GitHub y un subconjunto está disponible en Kaggle, lo que facilita su acceso y uso por parte de investigadores y desarrolladores.

En conjunto, el conjunto de datos CPPE-5 tiene el potencial de impulsar significativamente la investigación en detección de objetos para equipos de protección personal médico. Se espera que este conjunto de datos sea utilizado para desarrollar algoritmos de detección de objetos más precisos y robustos que puedan ayudar a mejorar la seguridad y la salud en entornos médicos y otros entornos públicos.

Además de las contribuciones mencionadas anteriormente, el artículo que presenta el conjunto de datos CPPE-5 también proporciona:

- Una revisión exhaustiva de los trabajos relacionados en detección de objetos y conjuntos de datos.
- Una descripción detallada del proceso de recopilación y anotación del conjunto de datos CPPE-5.
- Un análisis estadístico del conjunto de datos.
- Resultados experimentales que comparan el rendimiento de diferentes modelos de detección de objetos en el conjunto de datos CPPE-5.

En general, el artículo presenta una contribución significativa al campo de la visión por computadora y la detección de objetos. El conjunto de datos CPPE-5 y los resultados asociados son valiosos para los investigadores y desarrolladores que trabajan en la detección de equipos de protección personal médico y otras aplicaciones de detección de objetos en el mundo real.