SN Computer Science (2023) 4:263
https://doi.org/10.1007/s42979-023-01748-7

ORIGINAL RESEARCH

CPPE-5: Medical Personal Protective Equipment Dataset

Rishit Dagli[1] · **Ali Mustufa Shaikh**[2]

Received: 13 December 2022 / Accepted: 22 February 2023 / Published online: 16 March 2023

Abstract
We present a new challenging dataset, CPPE-5 (Medical Personal Protective Equipment), with the goal to allow the study of subordinate categorization of medical personal protective equipments, which is not possible with other popular data sets that focus on broad-level categories (such as PASCAL VOC, ImageNet, Microsoft COCO, OpenImages, etc). To make it easy for models trained on this dataset to be used in practical scenarios in complex scenes, our dataset mainly contains images that show complex scenes with several objects in each scene in their natural context. The image collection for this dataset focuses on: obtaining as many non-iconic images as possible and making sure all the images are real-life images, unlike other existing datasets in this area. Our dataset includes five object categories (coveralls, face shields, gloves, masks, and goggles), and each image is annotated with a set of bounding boxes and positive labels. We present a detailed analysis of the dataset in comparison to other popular broad-category datasets as well as datasets focusing on personal protective equipments, we also find that at present, there exist no such publicly available datasets. Finally, we also analyze performance and compare model complexities on baseline and state-of-the-art models for bounding box results. Our code, data, and trained models are available at https://git.io/cppe5-dataset.

Keywords Ground-truth dataset · Computer vision · Object detection

Introduction

Deep learning is revolutionizing multiple areas of computer vision. An explosive popularity in this field was brought after the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [1] and has pushed forward the state of the art in generic object detection. It contains a detection challenge using ImageNet images [2]. Since then, the performance of models has been improving at unparalleled speeds. Among the many challenges in machine learning, data collection is becoming one of the critical bottlenecks [3]. As deep learning becomes popular the core of their success is the need for rich and large annotated training data [4]. Larger and richer annotated datasets are a boon for leading-edge research in computer vision to enable the next generation of state-of-the-art algorithms [5] and have been instrumental in driving progress in object recognition over the last decade [6–9].

Object detection is a fundamental problem of computer vision that deals with detecting instances of visual objects of a certain class in digital images. The objective of object detection aims to develop models and techniques to provide the information: "what objects are where?" [10] Building larger and richer datasets often play a key role in allowing computers to identify and interpret images as compositions of one or multiple objects which has been quite tricky for machines so far [11]. Through this object detection dataset, we majorly aim to advance machines to automatically identify where objects (personal protective equipments) are precisely located.

In object detection, a number of well-known datasets and benchmarks have been released in the past 10 years. Most datasets contain a wide variety of common-level classes, such as different kinds of animals or inanimate things. Several such datasets have emerged as standards for the community including MIT-CSAIL [12], PASCAL VOC Challenges (e.g., VOC2007, VOC2012) [13, 14], ImageNet [2],

✉ Rishit Dagli
rishit.dagli@mail.utoronto.ca

Ali Mustufa Shaikh
ali.shaikh@postman.com

[1] Department of Computer Science, University of Toronto, 40 St. George Street, Toronto, ON M5S3H4, Canada

[2] Postman Inc., 309, Venkatesh complex, Bangalore, Karnataka 560038, India

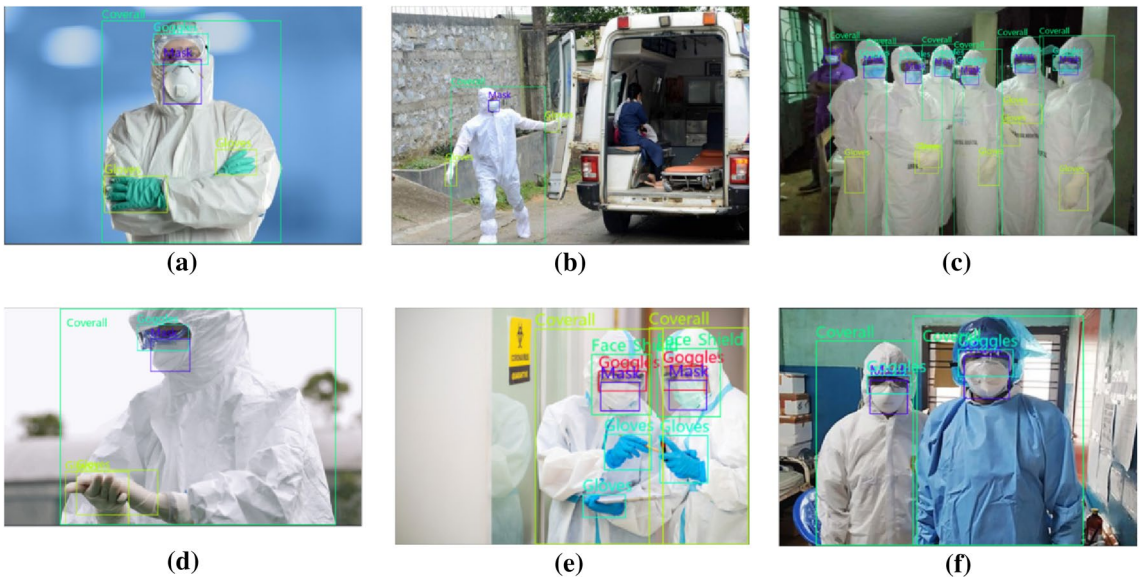

Fig. 1 Example annotations in CPPE-5 for object detection demonstrating the five classes of our data set. Each example image is shown with an outline (bounding box) and the object it is identified as

Caltech-256 [15], Microsoft COCO [16] and DOTA [17, 18]. However, this dataset was built bearing in mind to allow for subordinate categorization especially for detecting personal protective equipment which is not possible with other large-scale popular datasets that focus on rather broad categories.

Though the first part subset of the dataset was released to facilitate working on Medical Personal Protective Equipments, these were carefully ported to create the final dataset expanding the goals to medical personal protective equipments. COVID-19 is causing widespread morbidity and mortality globally. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) responsible for this disease infected more than 17 million people by August 2020 [19]. It has also been observed that the global trend is approximately exponential, at a rate of tenfold every 19 days [20]. Considering this, it is very important to be able to accurately detect Medical Personal Protective Equipment to help limit the growth of COVID-19. To encourage the development of such tools we present this dataset publicly on GitHub and a subset of the dataset on Kaggle[1] focusing on accurately identifying the Personal Protective Equipments through images.

In this paper, we introduce the CPPE-5 (Medical Personal Protective Equipment), an object detection dataset, which contains images and ground-truth annotations for the task of object detection. The majority of the images have been collected from Flickr,[2] with an aim to collect a majority of non-iconic images. A small portion of images was collected from Google Images as well. After doing so each of the images was annotated using crowd-sourcing techniques [21]. Each of these annotations were evaluated by multiple people and were also then evaluated by us to keep a strict check on the quality of the ground-truth annotations.

As mentioned, we provide unified annotations for the task of object detection with the dataset. In Fig. 1, we show examples of annotations provided in the dataset. In (a) coveralls, gloves, mask, goggles; (b) coveralls, gloves, mask; (c) coveralls, gloves, goggles; (d) coveralls, gloves, mask, goggles; (e) coveralls, gloves, mask, goggles, face shield; and (f) coveralls, mask, goggles are demonstrated. Some more sample images for each category could be found in Appendix 2.

With the CPPE-5 dataset, we hope to facilitate research and use in applications at multiple public places to autonomously identify if a PPE kit has been worn and also which part of the PPE kit has been worn. One of the main aims of this dataset was to also capture a higher ratio of non-iconic images or non-canonical perspectives [22] of the objects in this dataset. We further hope to see high use of this dataset to aid in medical scenarios which would have a huge effect worldwide.

The remainder of this article is organized as follows: In "Related Work", related works are given. In "Dataset Collection and Annotation',' we describe the process used to

[1] https://www.kaggle.com/ialimustufa/object-detection-for-ppe-covid19-dataset.

[2] https://www.flickr.com/.

collect and annotate the dataset. In "Dataset Statistics", we present statistics related to the dataset. In "Experimental Results", we present the experimental results, training multiple state-of-the-art and baseline models. In "Conclusion", we conclude the article and give future works.

Related Work

Throughout the history of computer vision research rich and large datasets have played a very important role. They not only provide a means to train and evaluate algorithms, but they also drive research in new and more challenging directions [16]. Earlier datasets like the Caltech-256 Object Category Dataset [15] and the MIT Pedestrian Database [23] facilitated the direct comparison of hundreds of computer vision algorithms and also pushed toward more complex problems. Recent datasets like The Open Images dataset v4 with ~ 9.2 M images [5] ImageNet dataset [2] with ~ 14 M images and Microsoft COCO with ~ 2.5 M labeled instances [16] have enabled breakthroughs in object detection research with a new wave of deep-learning algorithms.

Performing object detection often requires identifying which specific class the object belongs to and also localizing the object in the image usually done with a bounding box as shown in Fig. 1. One of the earliest algorithms focused on face detection often using ad hoc datasets [24]. Later, more realistic and challenging datasets were built which facilitated the creation of many deep-learning algorithms. Transformers [25] were first introduced to vision in Vision Transformer (ViT) [26] by splitting an image into a sequence of visual tokens. The self-attention strategy in ViTs has demonstrated superior performance to modern convolutional neural networks (ConvNets) when trained with optimized recipes. A lot of popularity in using Transformers for object detection tasks was brought through DEtection TRansformer (DETR) [27] and achieved at-par results with earlier methods like Faster RCNN [28]. After this multiple works tried training transformers for object detection mainly using ViTs directly for object detection [26] and Swin Transformers [29]. Recently, self-attention and transformer-based methods have shown a lot of promise for object detection and dominated the state-of-the-art for this task [30–34].

For the detection of basic object categories the PASCAL VOC datasets [13] were created which contained 20 object categories, over (11,000) images, and over (27,000) annotated objects using bounding boxes of which almost (7000) had detailed segmentations. Later, the ImageNet dataset was created [2] which included over 14 M images across 1000 object categories. The ImageNet large-scale visual recognition challenge facilitated the creation of many deep-learning algorithms, namely AlexNet [6], Inception v1 [35], VGGNet [36], ResNet [37] and more. Later, the Microsoft COCO: Common Objects In Context dataset [16] was created for the detection and segmentation of objects occurring in their natural context. This dataset aimed to find non-iconic images containing objects in their natural context. The COCO dataset consists of over (330,000) across 91 categories with 1.5 M object instances.

Machine Learning for Health is quite a popular field with quite a lot of research pertaining to Machine Learning for COVID-related topics [38–40]. Many prior works aim to solve a binary classification problem: often if a mask is worn or not; masks are one of the most widely used components of a personal protective equipment kit. The datasets acquired in these papers were in controlled environments or simulated images however to deploy these tools majorly requires them to be robust to multiple variations (eg. lighting conditions, terrain, and background objects). In the next part of this section, we talk about some related work about identifying masks in images, masks being one of the most widely used objects and are also present in our dataset. However, to the best of our knowledge, we found no related work for the rest of the categories in our dataset.

Chowdary et al. [41] in their paper transfer learn on top of Inception V3 pre-trained on ImageNet dataset [42] for the task of binary classification: identifying if a mask has been worn or not. This paper also claims to achieve quite plausible results in testing on simulated data. However, the models proposed in this paper were trained and tested on simulated data: where an image of a mask was artificially superimposed later on top of the face images. Furthermore, the images in this dataset are all iconic face images on top of which a mask was artificially added, this tends to lose out not only on a lot of contextual information but models trained on this data are unable to identify all kinds of mask and masks worn in different positions due to the artificial training data. To this end, in our dataset, we have ensured each image is a real image and no objects were artificially added on the image. Our dataset also focuses on more than one category of personal protective equipment unlike this dataset which focuses on only masks.

Wang et al. [43] in their paper introduce three datasets Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD), and Simulated Masked Face Recognition Dataset (SMFRD) for the task of binary classification. The multi-granularity masked face recognition model developed in this paper also claims to achieve (95)% accuracy on the Real-world Masked Face Recognition dataset. The Real-world Masked Face Recognition dataset includes (5000) pictures of 525 people wearing masks, and (90,000) images of the same 525 subjects without masks. However, the images in this dataset are not necessarily medical masks. As an example, this dataset also includes images with a scarf worn or sports helmets and masks under the category of people wearing masks. In our

Table 1 Categories in the CPPE-5 dataset

Coveralls	Coveralls are hospital gowns worn by medical professionals in order to provide a barrier between patient and professional, these usually cover most of the exposed skin surfaces of the professional medics
Mask	Mask prevents airborne transmission of infections between patients and/or treating personnel by blocking the movement of pathogens (primarily bacteria and viruses) shed in respiratory droplets and aerosols into and from the wearer's mouth and nose
Face shield	Face shield aims to protect the wearer's entire face (or part of it) from hazards such as flying objects and road debris, chemical splashes (in laboratories or in the industry), or potentially infectious materials (in medical and laboratory environments)
Gloves	Gloves are used during medical examinations and procedures to help prevent cross-contamination between caregivers and patients
Goggles	Goggles, or safety glasses, are forms of protective eyewear that usually enclose or protect the area surrounding the eye in order to prevent particulates, water or chemicals from striking the eyes

Medical Personal Protective Equipment (CPPE-5) dataset, as we later mention in "Dataset Collection and Annotation", all the images have been checked for quality and relevance.

Loey et al. [44] in their paper also use the three above-mentioned datasets: Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD), and Simulated Masked Face Recognition Dataset (SMFRD) to train a binary classifier. The model proposed in this paper uses ResNet-50 [37] as a feature extractor and then uses traditional machine learning algorithms for classification. In this setting, the paper reports quite plausible performance on the Real-world Masked Face Recognition Dataset. However, this paper measures the performance of models by training on the Real-world Masked Face Recognition Dataset but majorly tests their models on simulated mask images and not real-world mask images. Our Medical Personal Protective Equipment (CPPE-5) dataset ensures all images are real-life images.

Nath et al. [45] in their paper aim to build a system to verify the Personal Protective equipment compliance of a construction worker. They also present an in-house dataset Pictor-v3 in this paper which contains 774 annotated images collected with crowd-sourcing techniques and 698 annotated images collected through web mining. In one of their approaches in this paper where their algorithm simultaneously detects individual workers and verifies PPE compliance with a single convolutional neural network is reported to achieve 72.3% mean average precision (mAP) in real-world settings. However, their dataset only includes three categories: worker, hat, and vest out of which only two are protective equipment categories: hat and vest. These object categories are also not well suited for medical scenarios. Our Medical Personal Protective Equipment (CPPE-5) dataset contains 5 categories of personal protective equipment, all of which are well suited for medical purposes.

Dataset Collection and Annotation

This section describes how we decided on the categories and collected the images in the Medical Personal Protective Equipment - 5 dataset.

Object Categories

To create a dataset, we had to ensure the categories we choose from a representative set of all categories, be relevant to practical applications, and occur with high enough frequency to enable the collection of a large dataset. A small group of daily Medical Personal Protective Equipment users were asked to share components of a PPE kit based on how often they are used and their usefulness for practical applications. Through this, we received seven potential categories for this dataset: coveralls or gowns, masks, face shields, gloves, shoe covers, respirators, and goggles

Some common PPE objects which are quite similar to the above list like lab coats, safety boots, full facepiece respirators, self-contained breathing apparatus, etc. were not included in the initial list of potential categories. Also, we omitted some PPE objects which are not used for medical scenarios from the initial categories; like helmets, harnesses, hearing protection, ballistic vests, etc. to maintain the focus of this dataset.

The final selection of categories attempts to pick categories for which obtaining a large number of images with categories in them was available. The final categories based on this did not include respirator and shoe cover due to a lack of rich annotations and enough data for these categories. The final object categories are denoted in the dataset as:

- Coveralls
- Face_Shield
- Gloves
- Goggles

Fig. 2 Example of **a** non-iconic images and **b** the little number of iconic images from our dataset

- Mask

We also show in detail about the categories in this dataset in Table 1. The category definitions shown in Table 1 were adapted from their Wikipedia[3] pages and were also used while annotating the datasets as shown in later sections.

Image Collection

Having decided on the object categories our next goal was to collect a set of candidate images. We classify images into two categories: iconic object images and non-iconic images as shown in Fig. 2. While iconic images (Fig. 2b): which have a single large object in a canonical perspective usually contain high-quality object instances they can lack important contextual information, these could be found directly by searching for the object category on Google Images[4] or Bing Image Search.[5] It has been shown by Torralba et al. [46] that non-iconic images are better at generalizing. We thus aimed to collect a majority of non-iconic images (Fig. 2a). This allows us to have a majority of complex images which contain several other objects.

As popularized by Caltech-UCSD Birds-200 [47, 48], Microsoft COCO [16] and Open Images v4 [5] datasets we majorly collected images from Flickr which tend to have lesser iconic images. Flickr contains images uploaded by millions of photographers with searchable metadata. A smaller portion of images was also collected from Google Images. We also remove near-duplicate images in the dataset using GIST descriptors [49, 50] greatly minimizing the chances of near-duplicate images in the dataset.

The images in the CPPE-5 dataset were collected using the following process:

Obtain Images from Flickr: Following the object categories we identified earlier, we first download images from Flickr and save them at the "Original" size. On Flickr, images are served at multiple different sizes (Square 75, Small 240, Large 1024, X-Large 4K, etc.), the "Original" size is an exact copy of the image uploaded by the author. In "Dataset Statistics", we talk more about the variation in image sizes and present statistics for the sizes of images in this dataset.

Extract relevant metadata: Flickr contains images each with searchable metadata, we extract the following relevant metadata:

- A direct link to the original image on Flickr
- Width and height of the image
- Title given to the image by the author
- Date and time the image was uploaded on
- Flickr username of the author of the image
- Flickr Name of the author of the image
- Flickr profile of the author of the image
- The License image is licensed under
- MD5 hash of the original image

Obtain Images from Google Images: Due to the reasons we mentioned earlier, we only collect a very small proportion of images from Google Images. For this set of images, we extract the following metadata:

- A direct link to the original image
- Width and height of the image
- MD5 hash of the original image

Filter inappropriate images: Though very rare in the collected images, we also remove images containing inappropriate content using the safety filters on Flickr and Google Safe Search.

Filter near-similar images: We then remove near-duplicate images in the dataset using GIST descriptors [51].

[3] https://www.wikipedia.org/.

[4] https://images.google.com/.

[5] https://www.bing.com/images.

Table 2 Frequency of the categories appearing in the CPPE-5 dataset calculated by the percentage of bounding boxes

Category	Coverall	Mask	Goggles	Face_Shield	Gloves
Frequency	25.48%	27.76%	8.66%	9.51%	28.59%

Image Annotation

In this section, we describe how we annotated our image collection. The dataset was labeled in two phases: the first phase included labeling 416 images and the second phase included labeling 613 images. In both phases, we used crowd-sourcing techniques with multiple volunteers labeling the dataset using the open-source tool LabelImg.[6] For all the images in the dataset volunteers were provided Table 1 as well as examples of correctly labeled images, incorrectly labeled images, and not applicable images. Before the labeling task, each volunteer was provided with an exercise to verify if the volunteer was able to correctly identify categories as well as identify if an annotated image is correctly labeled, incorrectly labeled, or not applicable.

The labeling process first involved two volunteers independently labeling an image from the dataset. In any of the cases where the number of bounding boxes is different, the labels for on or more of the bounding boxes are different or two volunteer annotations are sufficiently different; a third volunteer compiles the result from the two annotations to come up with a correctly labeled image. After this step, a volunteer verifies the bounding box annotations. Following this method of labeling, the dataset we ensured that all images were labeled accurately and contained exhaustive annotations. As a result of this, our dataset consists of 1029 high-quality, majorly non-iconic, and accurately annotated images.

In Table 2 we show the frequency of the categories in the Medical Personal Protective Equipment (CPPE-5) dataset. `Gloves` and `Mask` are the most common annotations, with a considerable portion of the bounding boxes being marked as such.

Dataset Statistics

Next, we analyze the properties of the Medical Personal Protective Equipment (CPPE-5) dataset. The Medical Personal Protective Equipment (CPPE-5) dataset contains 1029 images and 4698 object annotations consisting of 1343 glove annotations, 1304 mask annotations, 1197 coverall annotations, 447 face shield annotations, and 407 goggle annotations as shown in Table 3. Table 3 also includes the number of images that contain at least 1 annotation belonging to a specific category.

Table 3 Number of annotations in the dataset

Category	No. of annotations ≥ 1	No. of images with category annotation	Average annotations/image
Coverall	1197	799	1.50
Mask	1304	898	1.45
Goggles	407	312	1.30
Face_Shield	447	344	1.30
Gloves	1343	575	2.34
Total	4698		4.57

We also compare the goals of the Medical Personal Protective Equipment (CPPE-5) dataset with other previous object detection datasets namely ImageNet [2], PASCAL VOC 2012 [52], Microsoft COCO [16] and RMFD [43]. ImageNet's goals include capturing a large number of object categories, many of which are fine grained. PASCAL VOC's goals include object detection in natural images. Microsft COCO is designed for the detection of objects occurring in their natural context. Real-world Masked Face Recognition Dataset aims to detect masked faces. The Medical Personal Protective Equipment (CPPE-5) dataset is designed for subordinate object detection for Personal Protective Equipment.

Next, we present statistics for images present in the dataset. On average our dataset contains 4.57 annotations per image. As shown in Fig. 3a, we calculate the distribution of aspect ratios as measured by $(\frac{\text{width}}{\text{height}})$. Our dataset has an average aspect ratio of 1.40. We also measure the distribution of image sizes as measured by $(\sqrt{\text{width} \times \text{height}})$. Generally smaller objects are harder to recognize and require more contextual reasoning to recognize, our dataset has an average image size of 946.94 pixels.

Experimental Results

In this section, we evaluate baseline and state-of-the-art object detection models trained on the Medical Personal Protective Equipment Dataset (CPPE-5) through extensive experiments. In "Experimental Setup", we detail the experimental setup used. In "Baseline Models", we describe how we chose the baseline models and share the results of the baseline models. In "Evaluating State-of-the-Art Models", we present results for State-of-the-Art object detection techniques trained on the Medical Personal Protective Equipment Dataset (CPPE-5) and make inferences about the difficulty of the dataset. To foster easy reproducibility of the results

[6] https://github.com/tzutalin/labelImg.

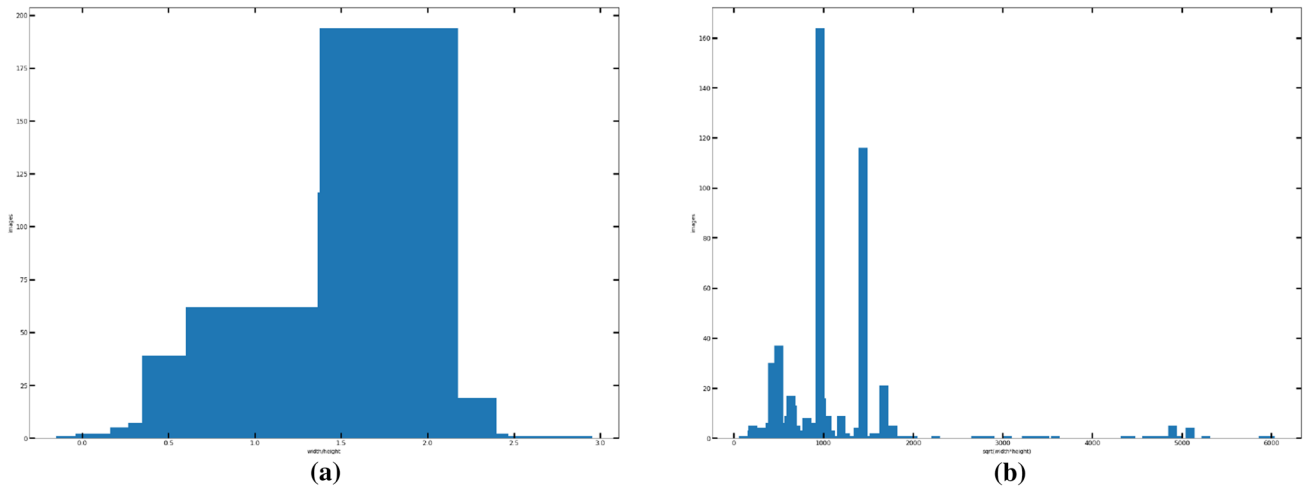

Fig. 3 Image statistics, **a** distribution of aspect ratios as measured by $(\frac{\text{width}}{\text{height}})$ and **b** distribution of image sizes as measured by $(\sqrt{\text{width} \times \text{height}})$. Overall the average aspect ratio is 1.40 and the image size is 946.94 pixels

Table 4 Baseline models trained on the CPPE-5 dataset

Method	AP^{box}	AP^{box}_{50}	AP^{box}_{75}	AP^{box}_{S}	AP^{box}_{M}	AP^{box}_{L}	#Params	Epochs
SSD [56]	29.50	57.0	24.9	32.1	23.1	34.6	64.34 M	160
YOLO [57]	38.5	79.4	35.3	23.1	28.4	49.0	61.55 M	273
Faster RCNN [28]	44.0	73.8	47.8	30.0	34.7	52.5	60.14 M	24

we present in this section, we have open-sourced the training code, trained models as well as the training logs as TensorBoard [53] dashboards in the associated code repository.

Experimental Setup

Our experiments are based on the open-source detection toolbox MMDetection [54] and implementations from the TensorFlow Model Garden [55]. The training is conducted on the 1029 training images and the models are tested using another set of 100 testing images. Depending on the throughput, the models were trained either on 8 Tesla A100 GPUs or on a Cloud TPUv3 cluster.

For evaluation, we adopt the metrics from the COCO detection evaluation criteria, including the mean Average Precision (AP) across IoU thresholds ranging from 0.50 to 0.95 at different scales which are standard for object detection tasks. The inference speed FPS (Frames per second) for the detector is measured on a machine with 1 Tesla V100 GPU.

Baseline Models

A significant gain was obtained in object detection with the introduction of Regions with CNN features (RCNN). DNNs, or the most representative CNNs, act in a quite different way from traditional approaches. They have deeper architectures with the capacity to learn more complex features than shallow ones. RCNN [58] brought the advances in image classification using deep learning to object detection using a two-stage approach: classify object proposal boxes into any of the classes of interest (Table 4).

Since the proposal of RCNN, a lot of improved models have been suggested, including Fast RCNN which jointly optimizes classification and bounding box regression tasks, and Faster RCNN which takes an additional sub-network to generate region proposals. Faster RCNN stills provide very competitive results today in terms of accuracy. More recently, single-shot detectors were presented to bypass the computational bottleneck of object proposals by regressing object locations directly from a predefined set of anchor boxes (e.g., SSD [56] and YOLO [57]). This typically results in simpler models that are easier to train end-to-end [5, 59]. All of them bring different degrees of detection performance improvements over the primary RCNN and make real-time and accurate object detection become more achievable.

We carefully choose Faster RCNN [28], YOLOv3 [57] and SSD [56] as our baseline testing algorithms for their excellent performance on general object detection. In Table 5, we present the results for these three baseline models.

Table 5 Top performing models based on the standard metric, box AP, trained on the CPPE-5 dataset

Method	(AP^{box})	(AP^{box}_{50})	(AP^{box}_{75})	(AP^{box}_{S})	(AP^{box}_{M})	(AP^{box}_{L})	#Params	Epochs
RepPoints [60]	43.0	75.9	40.1	27.3	36.7	48.0	36.6 M	24
Sparse RCNN [61]	44.0	69.6	44.6	30.0	30.6	54.7	124.99 M	36
FCOS [62]	44.4	79.5	45.9	36.7	39.2	51.7	50.8 M	24
Grid RCNN [63, 64]	47.5	77.9	50.6	43.4	37.2	54.4	121.98 M	25
Deformable DETR [65]	48.0	76.9	52.8	36.4	35.2	53.9	40.5 M	50
FSAF [66]	49.2	84.7	48.2	45.3	39.6	56.7	93.75 M	12
Localization distillation [67]	50.9	76.5	58.8	45.8	43.0	59.4	32.05 M	12
VarifocalNet [68]	51.0	82.6	56.7	39.0	42.1	58.8	53.54 M	24
RegNet [69]	51.3	85.3	51.8	35.7	41.1	60.5	31.5 M	24
Double heads [70]	52.0	87.3	55.2	38.6	41.0	60.8	148.7 M	12
DCN [71, 72]	51.6	87.1	55.9	36.3	41.4	61.3	148.71 M	12
Empirical attention [73]	52.5	86.5	54.1	38.7	43.4	61.0	47.63 M	12
TridentNet [74]	52.9	85.1	58.3	42.6	41.3	62.6	32.8 M	36

Faster RCNN was trained with a ResNet 101 backbone. Only random flip data augmentations were applied to the image. We use the SGD optimizer with a momentum of 0.9 and a weight decay of 0.0001 and no gradient clipping. We use a step learning rate scheduler with an initial learning rate of 0.02 with a linear warm-up for 500 iterations. We use sigmoid cross entropy loss as the classifier loss and L1 loss as the bounding box loss. This baseline model was trained for 24 epochs.

YOLO was trained with a DarkNet 53 backbone The data augmentation pipeline uses random flip, photometric distortion, and a random crop on the image and bounding boxes such that the cropped patches have minimum IoU requirement with the original image and bounding boxes. We use the SGD optimizer with a momentum of 0.9 and a weight decay of 0.0005 and apply gradient clipping using the L^2 norm. We use a step learning rate scheduled with an initial learning rate of 0.001 with a linear warm-up for 2000 iterations. We use sigmoid cross entropy loss as the classifier loss, confidence loss, and the *xy*-coordinate loss, and MSE loss for wh-coordinate loss. The *xy*-coordinate loss and wh-coordinate loss use a weight of 2. This baseline model is trained for 273 epochs.

SSD was trained with a MobileNet V1 backbone. Only random flip data augmentations were applied to the image. We use the momentum optimizer with a momentum of 0.9. We use a cosine decay learning rate schedule with an initial learning rate of 0.04 and warm-up for 2000 iterations with the learning rate $\frac{4}{300}$. We use weighted smoothed L^1 as the localization loss and weighted sigmoid focal as the classification loss with $\alpha = 0.25$ and $\gamma = 2.0$. This baseline model is trained for 160 epochs.

Evaluating State-of-the-Art Models

We also present results from training some state-of-the-art object detection models on Medical Personal Protective Equipment Dataset (CPPE-5) in Table 5 using the same evaluation procedure as mentioned earlier. Comparing these results with that of some other widely used object detection datasets like OpenImages, Microsoft COCO, and Pascal VOC,[7] we conclude that Medical Personal Protective Equipment Dataset (CPPE-5) does include more difficult (non-iconic) images of objects. We include more details on how each of these models was trained in the associated code repository.

[7] https://paperswithcode.com/sota.

Conclusion

This paper presented a new object detection dataset, the Medical Personal Protective Equipment Dataset (CPPE-5) which is the first dataset focusing on the subordinate category of medical Personal Protective Items and would have wide practical uses. We conducted a detailed analysis of the dataset and compared it to other popular broad-category datasets and datasets focusing on personal protective equipment. We found that there is currently no publicly available dataset for studying subordinate categorization of medical personal protective equipment. Overall, our CPPE-5 dataset fills a significant gap in the availability of datasets for the study of subordinate categorization of medical personal protective equipment. We annotate a huge number of well-distributed oriented objects with oriented bounding boxes with emphasis placed on finding non-iconic images of objects in natural environments and varied viewpoints. We assume this dataset is challenging but very similar to real-world scenarios, making this an appropriate dataset for practical applications. We explained how the data were collected and annotated and presented dataset statistics indicating that the images often contain multiple bounding boxes per image. We further also evaluated multiple modern state-of-the-art and baseline object detection models trained on our dataset, establishing a benchmark for subordinate categorization for medical Personal Protective Equipment images. Many object detection algorithms benefit from additional annotations, such as the amount an instance is occluded or the location of key points on the object which we believe are promising directions for future annotations. Detecting medical Personal Protective Equipments is a task of great practical importance, we believe CPPE-5 will not only promote the development of object detection algorithms for this purpose but also pose interesting algorithmic questions to general object detection in computer vision.

Appendix 1: Implementation Details

In this section, we explain the implementation details of the experiments we perform and the models we train.

Sampling There is a slight class imbalance in the dataset for some of the classes, meaning that not all classes have a similar number of images. For this reason, we follow a stratified sampling strategy during data loading.

Code Our code is in PyTorch 1.10 [75]. We use a number of open-source packages to develop our training workflows. Most of our experiments and models were trained with mmdetection [54] and we also used timm [76] for some of the experiments. We also utilized TensorFlow [77], TensorFlow Lite,[8] and TensorFlow.js[9] for creating edge deployment ready models for the mobile and browser. Furthermore, we also used Tensorboard [53] while training the model. Our hardware setup for the experiments included either eight NVIDIA Tesla A100 GPUs or a TPUv3 cluster. We utilized mixed-precision training with PyTorch's native AMP (through `torch.cuda.amp`) for mixed-precision training and a distributed training setup (through `torch.distributed.launch`) which allowed us to obtain significant boosts in the overall model training time.

Hyperparameters Due to the extent of our experiments, we redirect the reader to our GitHub repository to find the hyperparameters and configurations for each of the experiments in this paper.

Appendix 2: Sample Images

In Fig. 4, we show 8 sample images from each of the categories in the dataset with the object annotations superimposed on the images. It is noteworthy to know that some of the images may not have the original image sizes since the class names were superimposed on the image and we did not want the class names to be cut off. These visualizations were generated with FiftyOne [78].

[8] https://github.com/tensorflow/tensorflow/tree/master/tensorflow/lite.

[9] https://github.com/tensorflow/tfjs.

Fig. 4 Samples of annotated images for each category in the CPPE-5 dataset

Appendix 3: Comparing Model Complexities

In Table 6, we compare model complexities and their performance on the CPPE-5 (Medical Personal Protective Equipment) dataset. We measure model complexity in terms of the number of parameters of the model and FLOPs required to run a single instance of the model. In Figs. 5 and 6, we show a visual representation of comparing the model complexities.

Table 6 Comparison between model complexity, in terms of number of parameters (in millions), FLOPs (in billions), and frames per second on a Tesla V100 GPU, and (AP^{box})

Method	(AP^{box})	#Params	FLOPs	FPS
SSD	29.5	64.34 M	103.216 G	25.6
YOLO	38.5	61.55 M	193.93 G	48.1
RepPoints	43.0	36.6 M	189.83 G	18.8
Faster RCNN	44.0	60.14 M	282.75 G	15.6
Sparse RCNN	44.0	124.99 M	241.53 G	21.7
FCOS	44.4	50.8 M	272.93 G	9.7
Grid RCNN	47.5	121.98 M	553.44 G	7.7
Deformable DETR	48.0	40.5 M	195.47 G	18.8
FSAF	49.2	93.75 M	435.88 G	5.6
Localization distillation	50.9	32.05 M	204.71 G	19.5
VarifocalNet	51.0	53.54 M	180.05 G	4.8
RegNet	51.3	31.5 M	183.29 G	18.2
Double heads	52.0	148.7 M	220.05 G	9.5
DCN	51.6	148.71 M	219.97 G	16,6
Empirical attention	52.5	47.63 M	185.83 G	12.7
TridentNet	52.9	32.8 M	822.13 G	4.2

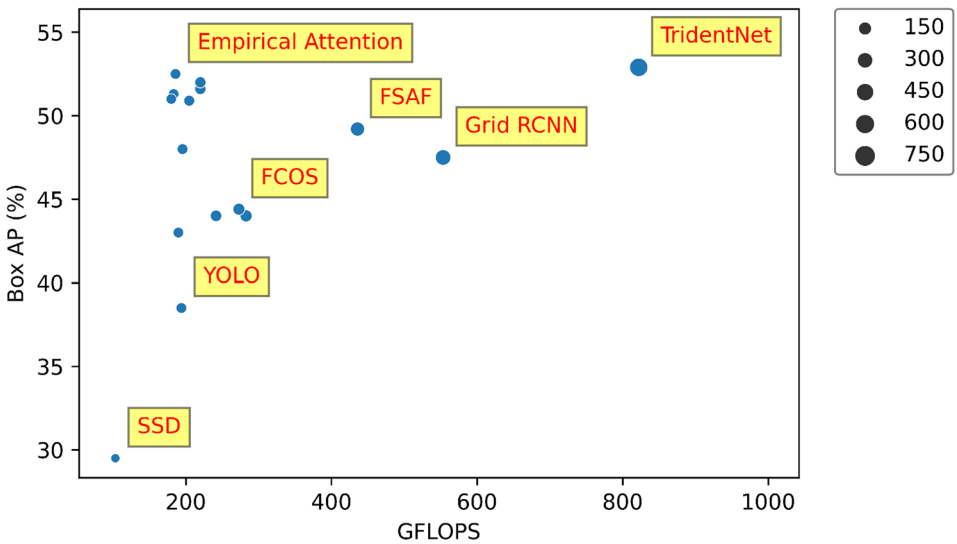

Fig. 5 Comparison of model performance and model complexity in terms of FLOPs (in billions)

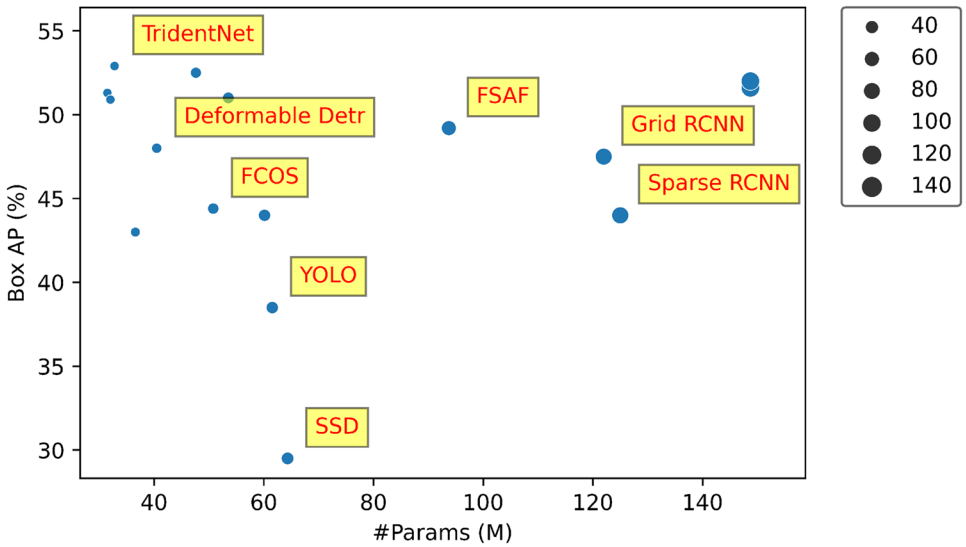

Fig. 6 Comparison of model performance and model complexity in terms of number of parameters (in millions)

Acknowledgements The authors would like to thank Google for supporting this work by providing Google Cloud credits. The authors would also like to thank Google TPU Research Cloud (TRC) program (https://sites.research.google/trc) for providing access to TPUs. The authors are also grateful to Omkar Agrawal for his help with verifying the difficult annotations.

Data Availability The datasets generated during and/or analyzed during the current study are available in the CPPE-5 repository, at https://git.io/cppe5-dataset.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L. Imagenet large scale visual recognition challenge. Int J Comput Vis. 2015;115(3):211–52. https://doi.org/10.1007/s11263-015-0816-y.
2. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee; 2009. p. 248–55.
3. Roh Y, Heo G, Whang SE. A survey on data collection for machine learning: a big data—AI Integration Perspective; 2019.
4. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press, Cambridge. 2016. http://www.deeplearningbook.org
5. Kuznetsova A, Rom H, Alldrin N, Uijlings J, Krasin I, Pont-Tuset J, Kamali S, Popov S, Malloci M, Kolesnikov A, Duerig T, Ferrari V. The open images dataset v4. Int J Comput Vis. 2020;128(7):1956–81. https://doi.org/10.1007/s11263-020-01316-z.
6. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1097–105.
7. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. OverFeat: integrated recognition, localization and detection using convolutional networks. 2014.
8. Viola P, Jones M, et al. Robust real-time object detection. Int J Comput Vis. 2001;4(34–47):4.
9. 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; 2016. p. 779–88.
10. Zou Z, Shi Z, Guo Y, Ye J. Object detection in 20 years: a survey. 2019. arXiv preprint arXiv:1905.05055.
11. Geirhos R, Janssen DHJ, Schütt HH, Rauber J, Bethge M, Wichmann FA. Comparing deep neural networks against humans: object recognition when the signal gets weaker. 2018.
12. Torralba A, Murphy KP, Freeman WT. Sharing features: efficient boosting procedures for multiclass object detection. In: Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition. CVPR 2004., vol. 2. IEEE. 2004. p. 2004.
13. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. Int J Comput Vis. 2010;88(2):303–38.
14. Everingham M, Eslami SA, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes challenge: a retrospective. Int J Comput Vis. 2015;111(1):98–136.
15. Griffin G, Holub A, Perona P. Caltech-256 object category dataset. California Institute of Technology. 2007.
16. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft coco: common objects in context. In: European conference on computer vision. Springer; 2014. p. 740–55.
17. Xia G-S, Bai X, Ding J, Zhu Z, Belongie S, Luo J, Datcu M, Pelillo M, Zhang L. Dota: a large-scale dataset for object detection in aerial images. In: The IEEE conference on computer vision and pattern recognition (CVPR). 2018.
18. Ding J, Xue N, Xia G-S, Bai X, Yang W, Yang MY, Belongie S, Luo J, Datcu M, Pelillo M, Zhang L. Object detection in aerial images: a large-scale benchmark and challenges. 2021.
19. Merow C, Urban MC. Seasonality and uncertainty in global covid-19 growth rates. Proc Natl Acad Sci. 2020;117(44):27456–64.
20. Li Y, Liang M, Yin X, Liu X, Hao M, Hu Z, Wang Y, Jin L. Covid-19 epidemic outside china: 34 founders and exponential growth. J Investig Med. 2021;69(1):52–5.

21. Vaughan JW. Making better use of the crowd: how crowdsourcing can advance machine learning research. J Mach Learn Res. 2017;18(1):7026–71.
22. Cutzu F, Edelman S. Canonical views in object representation and recognition. Vis Res. 1994;34(22):3037–56. https://doi.org/10.1016/0042-6989(94)90277-1.
23. Papageorgiou C, Poggio T. A trainable system for object detection. Int J Comput Vis. 2000;38(1):15–33.
24. Hjelmås E, Low BK. Face detection: a survey. Comput Vis Image Underst. 2001;83(3):236–74.
25. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
26. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16 × 16 words: transformers for image recognition at scale. 2020. arXiv preprint arXiv:2010.11929.
27. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: Computer vision—ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part I 16. Springer. 2020. p. 213–229
28. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015;28.
29. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021. p. 10012–10022.
30. Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, Ning J, Cao Y, Zhang Z, Dong L, Wei F, Guo B. Swin transformer v2: scaling up capacity and resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2022. p. 12009–12019.
31. Zhang H, Li F, Liu S, Zhang L, Su H, Zhu J, Ni LM, Shum H-Y. DINO: DETR with improved DeNoising anchor boxes for end-to-end object detection. arXiv (2022). https://doi.org/10.48550/ARXIV.2203.03605.
32. Wei Y, Hu H, Xie Z, Zhang Z, Cao Y, Bao J, Chen D, Guo B. Contrastive learning rivals masked image modeling in fine-tuning via feature distillation. arXiv 2022. https://doi.org/10.48550/ARXIV.2205.14141.
33. Chen Q, Wang J Han C, Zhang S Li, Z, Chen X, Chen J, Wang X, Han S, Zhang G, Feng H, Yao K, Han J, Ding E, Wang J. Group DETR v2: strong object detector with encoder-decoder pretraining. arXiv 2022. https://doi.org/10.48550/ARXIV.2211.03594.
34. Zong Z, Song G, Liu Y. DETRs with collaborative hybrid assignments training. arXiv 2022. https://doi.org/10.48550/ARXIV.2211.12860.
35. Szegedy, C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 1–9.
36. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2015.
37. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770–778.
38. Kushwaha S, Bahl S, Bagha AK, Parmar, KS, Javaid M, Haleem A, Singh RP. Significant applications of machine learning for covid-19 pandemic. J Ind Integr Manag. 2020;5(4).
39. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight covid-19. Physiol Genom. 2020;52(4):200–2. https://doi.org/10.1152/physiolgenomics.00029.2020. (**PMID: 32216577**).
40. Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT. New machine learning method for image-based diagnosis of covid-19. PLoS ONE. 2020;15(6):1–18. https://doi.org/10.1371/journal.pone.0235187.
41. Chowdary GJ, Punn NS, Sonbhadra SK, Agarwal S. Face mask detection using transfer learning of inceptionv3. In: International conference on big data analytics. Springer. 2020. p. 81–90.
42. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), 2016. p. 2818–26. https://doi.org/10.1109/CVPR.2016.308
43. Wang Z, Wang G, Huang B, Xiong Z, Hong Q, Wu H, Yi P, Jiang K, Wang N, Pei Y, Chen H, Miao Y, Huang Z, Liang J. Masked face recognition dataset and application. 2020.
44. Loey M, Manogaran G, Taha MHN, Khalifa NEM. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic. Measurement. 2021;167:108288. https://doi.org/10.1016/j.measurement.2020.108288.
45. Nath ND, Behzadan AH, Paal SG. Deep learning for site safety: real-time detection of personal protective equipment. Autom Constr. 2020;112:103085.
46. Torralba A, Efros AA. Unbiased look at dataset bias. In: CVPR 2011; 2011. p. 1521–1528. https://doi.org/10.1109/CVPR.2011.5995347.
47. Wah C, Branson S, Welinder P, Perona P, Belongie S. The caltech-ucsd birds-200-2011 dataset. California Institute of Technology. 2011.
48. Welinder P, Branson S, Mita T, Wah C, Schroff F, Belongie S, Perona P. Caltech-ucsd birds 200. California Institute of Technology. 2010.
49. Douze M, Jégou H, Sandhawalia H, Amsaleg L, Schmid C. Evaluation of gist descriptors for web-scale image search. In: Proceedings of the ACM international conference on image and video retrieval. CIVR '09. Association for Computing Machinery, New York. 2009. https://doi.org/10.1145/1646396.1646421
50. Murillo AC, Singh G, Kosecka J, Guerrero JJ. Localization in urban environments using a panoramic gist descriptor. IEEE Trans Rob. 2012;29(1):146–60.
51. Douze M, Jégou H, Sandhawalia H, Amsaleg L, Schmid C. Evaluation of gist descriptors for web-scale image search. In: Proceedings of the ACM international conference on image and video retrieval; 2009. p. 1–8.
52. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The PASCAL visual object classes challenge 2012 (VOC2012) results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html
53. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G.S, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow. Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org 2015. https://www.tensorflow.org/
54. Chen K, Wang J, Pang J, Cao Y, Xiong Y, Li X, Sun S, Feng W, Liu Z, Xu J, Zhang Z, Cheng, D, Zhu C, Cheng T, Zhao Q, Li B, Lu X, Zhu R, Wu Y, Dai J, Wang J, Shi J, Ouyang W, Loy CC, Lin D. MMDetection. Open mmlab detection toolbox and benchmark. 2019. arXiv preprint arXiv:1906.07155.
55. Yu H, Chen C, Du X, Li Y, Rashwan A, Hou L, Jin P, Yang F, Liu F, Kim J, Li J. TensorFlow model garden. 2020. https://github.com/tensorflow/models.

56. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. Ssd: single shot multibox detector. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer vision—ECCV 2016. Cham: Springer; 2016. p. 21–37.
57. Redmon J, Farhadi A. YOLOv3: an incremental improvement. 2018.
58. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 580–587.
59. Zhao Z-Q, Zheng P, Xu S-t, Wu X. Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst. 2019;30(11):3212–32.
60. Yang Z, Liu S, Hu H, Wang L, Lin S. Reppoints: point set representation for object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019. p. 9657–9666.
61. Sun P, Zhang R, Jiang Y, Kong T, Xu C, Zhan W, Tomizuka M, Li L, Yuan Z, Wang C, et al. Sparse r-cnn: end-to-end object detection with learnable proposals. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. p. 14454–14463.
62. Tian Z, Shen C, Chen H, He T. Fcos: fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019. p. 9627–9636.
63. Lu X, Li B, Yue Y, Li Q, Yan J. Grid r-cnn. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2019. p. 7355–7364. https://doi.org/10.1109/CVPR.2019.00754
64. Lu X, Li B, Yue Y, Li Q, Yan J. Grid R-CNN plus: faster and better. 2019.
65. Zhu X, Su W, Lu L, Li B, Wang X, Dai J. Deformable detr: Deformable transformers for end-to-end object detection. 2020. arXiv preprint arXiv:2010.04159.
66. Zhu C, He Y, Savvides M. Feature selective anchor-free module for single-shot object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 840–849.
67. Zheng Z, Ye R, Wang P, Wang J, Ren D, Zuo W. Localization distillation for object detection. 2021.
68. Zhang H, Wang Y, Dayoub F, Sunderhauf N. Varifocalnet: An iou-aware dense object detector. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2021. p. 8514–8523.
69. Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P. Designing network design spaces. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 10428–10436.
70. Wu Y, Chen Y, Yuan L, Liu Z, Wang L, Li H, Fu Y. Rethinking classification and localization for object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 10186–10195.
71. Dai J, Qi H Xiong Y, Li Y, Zhang G, Hu H, Wei Y. Deformable convolutional networks. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 764–773
72. Zhu X, Hu H Lin S Dai J. Deformable convnets v2: More deformable, better results. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 9308–9316.
73. Zhu X, Cheng D, Zhang Z, Lin S, Dai J. An empirical study of spatial attention mechanisms in deep networks. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV). 2019.
74. Li Y, Chen Y, Wang N, Zhang Z. Scale-aware trident networks for object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019. p. 6054–6063.
75. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. Pytorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 2019;32.
76. Wightman R. PyTorch image models github. 2019. https://doi.org/10.5281/zenodo.4414861.
77. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X. Tensorflow: A system for large-scale machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16); 2016. p. 265–283. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.
78. Moore BE, Corso JJ. Fiftyone. GitHub. Note: https://github.com/voxel51/fiftyone. 2020.