A Critical Analysis of Image Annotation Tools

Manisha Gupta
Department of Physics and Computer Science,
Faculty of Science,
Dayalbagh Educational Institute
Agra, UP, India

Dr. Preetvanti Singh
Department of Physics and Computer Science,
Faculty of Science,
Dayalbagh Educational Institute
Agra, UP, India

Abstract

Image annotation, a promising and relatively new area, is defined as an act of assigning descriptive text or categorizing the images by utilizing annotation tools, the features that the user wants to recognize in the model on its own. Image annotation is a crucial undertaking within the field of artificial intelligence, as it grants machines the visual perception and understanding to interpret the world. Research in the technology of image annotation continues to grow in various computer vision tasks over the coming decades. Several tools are developed to provide timely and efficient image labeling. The objective of this review paper is to provide a summary of commonly used image annotation tools. The paper first gives an overview of image annotation including the types and techniques of image annotations. Next, the paper describes the applications in various fields like agriculture, medical, environment, security and transportation. In next section, the paper gives the critical analysis of image annotation tools with their descriptions. Last section of the paper describes the capabilities of image annotation tools and the techniques supported by the chosen annotation tools, presents the formats of the images, a table for the utilization of the tools and comparison chart of the image annotation tools. Survey databases like Google Scholar, ACM Library, and IEEE Explorer are surveyed to generate statistics related to the tools. The selected image annotation tools are analyzed and classified using their own source.

Keywords Image annotation; image annotation tools; image annotation tools applications; image annotation techniques.

Introduction

A subfield of artificial intelligence (AI) called computer vision teaches computers how to capture and understand useful information gleaned from digital pictures, videos, and other visual sources. It enables the machine to accurately identify and locate objects, and then react to what they see. Every image in the dataset must be thoughtfully and accurately labeled to train an AI system to recognize objects like the ways humans can. Image annotation is one of the most important stages in the development of computer vision and image recognition applications. It involves recognizing, obtaining, describing, and interpreting results from digital images or videos.

Image annotation is the process by which a computer system assigns metadata in the form of captioning or keywords to a digital image. It is a new area to research that gives eyes to a machine. Training models are developed for predetermining the labels (or classes) to show the data features in the image. Figure 1 shows an example of an annotated image with identified labels or classes (such as car, person, bus, traffic light, and truck) in an image. These tagged images are then used to train the computer to identify those characteristics when presented with fresh and unlabeled data. More on image annotation can be found in Zhang, Islam & Lu (2012).

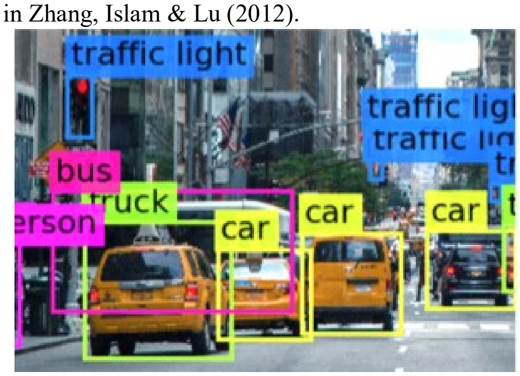

Fig. 1 Annotated Image
(*Source*: https://www.telusinternational.com/articles/)

To annotate images, several image annotation tools are available, capabilities of which can be applied to different research areas like medical imaging, drone imagery analysis, robotics, and agriculture. This paper presents an overview of image annotation and the tools used for annotating the images. These tools are characterized based on the covered tasks, and the provided functionalities. The objectives of this paper are to provide novice readers with a basic knowledge of image annotation and to critically analyze the usage of image annotation tools in various real-life applications.

OVERVIEW OF IMAGE ANNOTATION

Image annotation is used to recognize objects and boundaries in an image, or to segment the images. It marks the features in an image that a user wants the machine learning system to recognize. These images can then be used to train the model using supervised learning. Once the model is deployed, it can identify those features in images that have not been annotated and which can be finally used in making decisions or taking some actions.

Image annotation may be simple which means labeling an image with a phrase that describes the objects pictured in it. For example, annotating an image of a cat with the label *domestic house cat*. Image annotation may also be complex which means identifying, counting, or tracking multiple objects or areas in an image. An example of this is annotating the difference between breeds of cats.

The two things required to label an image are an image annotation tool and quality training data. Amongst the wide variety of image annotation tools, the need is to choose the tool that fits the use case of a user. This requires a deep understanding of the type of data that is being annotated and the task at hand. The focus must be on the modality of the data, the type of annotation required, and the format in which annotations are to be stored.

Annotations can be done on an individual or organizational level or can be outsourced to freelancers or organizations offering annotation services.

An image annotation tool is required to label the visual data. These tools may be available for a fee. There are open-source image annotation tools that can also be used freely. The open-source tools are modifiable, i.e., these can be modified according to the business needs of the users.

The process of the annotation works in the following steps:

1. **Preparing the image dataset:** The image dataset includes any type of image one wants to perform annotation on. Common types of data used with image annotation are 2D or 3D images/videos. This may include data from cameras or other imaging techniques such as an optical microscope (for 2D) or scanning probe microscopes (for 3D).

   To perform image annotation, it is not always required to input an image, it can also be done by importing images from standard datasets. Various datasets available for image annotation are Google's Open Images, MS COCO, LabelMe, and Visual QA. With the right image datasets, a data scientist can teach a computer to essentially function as though it had eyes of its own.

   There is no single standard format for image annotation. However, commonly used annotation formats are:
   - **COCO**, stores annotations using JSON.
   - **Pascal VOC,** stores annotation in an XML file.
   - **YOLO**, stores annotation in a .txt file.

2. **Finding out the Label types to be used:** Figuring out what type of annotation to use relates to the kind of task taught to the learning algorithm.

3. **Creating a class for each object to be Labelled:** Supervised learning algorithms must run on data that has a fixed number of classes.

4. **Annotate with the right tool:** To annotate the image, the shape in the image is selected and drawn around the object to be detected. The corresponding image region can be annotated depending on the computer vision task.

   While annotating the images, the following types of information can be associated with images:
   - *Content-independent metadata* is related to the image content, but it does not describe the content directly. For example, the author's name and location;

- *Content-dependent metadata* directly relates to the visual content of images and refers to low or intermediate-level features, for example, color, texture, and shape;
- *Content-descriptive metadata* directly relates to the visual content of images and refers to content semantics and can be specified using one or more of the following approaches:
  - *Free text descriptions*: No pre-defined structure for the annotation is given.
  - *Keywords*: Keywords to describe an image may be chosen arbitrarily from controlled vocabularies or from restricted vocabularies defined in advance.
  - *Classifications based on ontologies*: Ontologies, a large classification system that classifies different aspects of life into hierarchical categories, are used for annotation. This is similar to classification by keywords, but keywords belonging to a hierarchy enrich the annotations. For example, it can easily be found out that a *dog* is a subclass of the class *animal*.

5. **Exporting the dataset:** Once the image is annotated, data can be exported in different formats depending upon the way it is to be used. Popular methods for exporting are JSON, XML, and pickle. For training deep learning algorithms there are other formats like COCO and Pascal VOC.

By repeating this process with different images, a machine learning model can be trained to identify the objects in unlabeled images on its own.

Types of Image Annotation

Image annotation can be used for training the models to perform four primary tasks of computer vision. Image annotation distinctly reveals particular features or areas within the image for each of these tasks.

1. **Image Classification**

   Image classification is a type of image annotation that deals with identifying the *presence* of similar objects depicted in images across an entire dataset. Preparing images for image classification is referred to as *tagging*. The classification applies across an entire image at a high level. For example, interior images of a home may be tagged with labels such as the *kitchen* or *living room*.

2. **Object Recognition/Detection**

   Object recognition is a form of image annotation that deals with identifying the *presence, location,* and the *number* of one or more objects in an image and labeling them accurately. It also can be used to identify a single object. As an example, in the images of street scenes trucks, cars, bikes, and pedestrians can be labeled. These could be labeled separately in the same image.

3. **Segmentation**

   Segmentation is an image annotation type that is used to determine how objects within an image are the same or different. It also can be used to identify differences over time. These are of three types:

   i. **Semantic segmentation** delineates boundaries between similar objects and labels them under the same identification. This method is useful in understanding the *presence, location,* and sometimes, the *size* and *shape* of objects. For example, in the baseball game images which include both the stadium crowd and the playing field, the crowd could be annotated to segment the seating from the field.

   ii. **Instance segmentation** tracks and counts the *presence, location, count, size,* and *shape* of objects in an image. This type of image annotation is also referred to as *object class*. Taking the example of baseball game images, each individual could be labeled in the stadium to determine how many people were in the crowd. Either semantic or instance segmentation can be performed pixel-wise, which means every pixel inside the outline is labeled. This can also be performed with boundary segmentation, where only the border coordinates are counted.

   iii. **Panoptic segmentation** blends semantic and instance segmentation to provide data that is labeled for both background (semantic) and object (instance). For example, panoptic segmentation can be used with satellite imagery to detect changes in protected conservation areas.

4. **Boundary Recognition**

   This form of image annotation enables the recognition of lines or boundaries of objects in an image. Boundaries can include the edges of an individual object, areas of topography shown in an image, or man-made boundaries that are present in the image. For example, boundary recognition can be used to identify traffic lanes, land boundaries, or sidewalks.

Image Annotation Techniques

Training intelligent machines require precise datasets that are processed using different image annotation techniques. These techniques help in creating training data that reflects the diversity of the real world. Different techniques for annotating an image are:

1. **Bounding Box Annotation**

A bounding box, the most used annotation technique, is a rectangular box used for defining the location of the object in an image. These annotations are generally needed for object detection algorithms where the box denotes the object boundaries. These can also be used to train an algorithm for self-driving cars and intelligent video analytics.

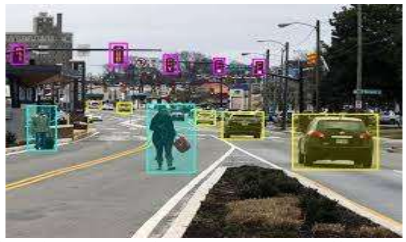

Fig. 2 Bounding box example (*Source: https://keymakr.com/blog/*)

A box is drawn around the object to be annotated within the image or these can be created by specifying the X and Y coordinates for the upper left and bottom right corners of the box. In the case of more than one target, objects may be chosen from a list of label attributes to the object within the box. In case a portion of the target object is blocked, the location of the target object's blocked edge(s) may be approximated.

2. **Polygons/Contours Annotation**

Polygons are used to annotate irregular objects within an image. Objects in an image are labeled by drawing an accurate contour around them. These are used to train object detection and semantic segmentation algorithm. Polygon annotation is used in annotating medical image data and in scene text recognition and localization.

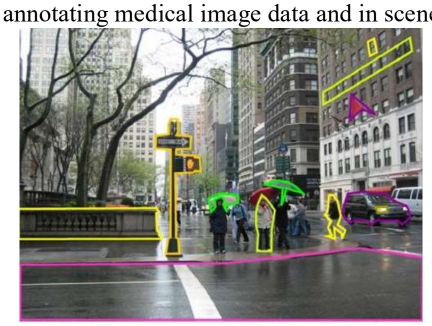

Fig. 3 Polygon example (*Source: https://oworkers.com/*)

3. **Lines Annotation**

Lines annotation involves the creation of lines and splines to delineate boundaries between one part of an image and another. This approach is useful for identifying boundaries to annotate road marks and sidewalks. This can also be used for trajectory planning in drones. This annotation type can be used to train algorithms for self-driving cars to choose lanes accurately.

Fig. 4 Lines and Polylines example *(Source: https://webtunix.ai/static/img)*

4. **Landmark Annotation**

Landmark annotation is also known as *dot annotation*. This involves creating dots or points across an image. Landmarking is used to identify fundamental points of interest within an image. Such points are referred to as *landmarks* or *key points*. Landmarking is significant in face recognition. This can also be used in gesture recognition and human pose recognition.

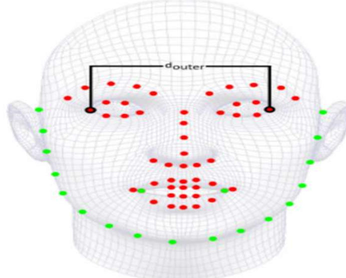

Fig. 5 Landmarking example
(*Source: https://www.researchgate.net/figure/ fig4_287211438*)

5. **Cuboid Annotation**
Cuboid Annotation creates a 3D representation of an object. It is similar to bounding boxes. A 3D bounding box (or cuboid) can also show an approximate depth of the target objects being annotated. To create an annotation, anchor points are placed at the edges of the item, and the space between the anchors is filled in with a line. These annotations are usually used in medical domains or training algorithms for the motion of robots and cars in a 3D environment.

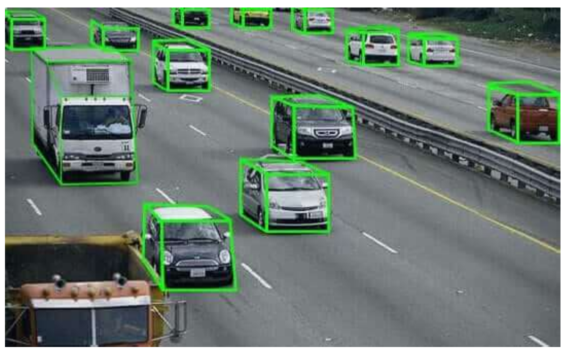

Fig. 6 Cuboid Annotation example
(*Source: https://www.cogitotech.com/3d-cuboid-annotation*)

6. **Semantic Segmentation Annotation**
This approach of image annotation involves separating an image into different regions and assigning a label to every pixel in an image. Regions of an image that carry different semantic meanings or definitions are considered separate from other regions. For example, one portion of the image could be *sky*, while another could be *grass*. This type of annotation comes in both 2D and 3D forms. The annotation is mainly utilized for self-driving cars and medical imaging.

Fig. 7 Semantic Segmentation Annotation example
(*Source: https://github.com/abreheret/*)

Review of the applications of image annotation shows that it is an essential step in computer vision.

CRITICAL ANALYSIS OF IMAGE ANNOTATION TOOLS

Image annotation tools provide an efficient method to develop machine learning models for annotating images. These tools can customize the analysis and result of the annotation that can be later used by the users in any image processing task.

Image annotation tools enable the user to label an object of interest in a frame by supporting manual, semi-automatic, and/or automatic labeling. There is a wide variety of image annotation tools that can be characterized based on the features and functionalities they support. Each of the image annotation tools has its annotation schema. This Section presents an overview of 27 image annotation tools released recently or used by a large group of people. The critical analysis of tools include describing the features and capabilities of each image annotation tool.

Description

- **Anno Mage (https://github.com/virajmavani/)**
  Anno-Mage is an image annotation software developed in Python. It is a freely available semi-automatic tool that suggests annotations for 80 object classes. It uses TensorFlow and karas object detection models. It incorporates the RetiNaNet model for object detection. The software uses COCO object classes for input, however, can be used with other datasets also. The software runs on Windows 10, Linux, and Mac operating systems.

- **Computer Vision Annotation Tool (https://cvat.org/)**
  CVAT (Computer Vision Annotation Tool) is an image annotation software written in TypeScript and Python. It is an open-source image annotation tool that can be used online or can also be installed on the local machine. CVAT is an interactive image annotation tool that supports the formats like COCO, Pascal VAC, YOLO, and many others. CVAT supports supervised machine learning tasks of object detection, image classification, and image segmentation. The software runs on Windows 10 and Linux operating systems.

- **DataTurks (https://github.com/DataTurks/)**
  DataTurks is an image annotation software developed in Java. It is freely available image annotation tool that can be used offline as well as in online mode. DataTurks uses machine learning models for image annotation that generates real-time reports. It uses AutoML and Human-in-the-Loop interactions for image classification and image segmentation. Images can be uploaded in a ZIP file. Export format can be Pascal VOC or TensorFlow. The software runs on Windows 10, Linux, and Mac operating systems.

- **Diffgram (https://diffgram.com/)**
  Diffgram is an image annotation software developed in Python. It is an open-source software that provides multiple tools in a single integrated application. The software supports segmentation and object detection by using machine learning and deep learning models, and ontologies. It has a simple user interface and can be easily installed. It supports .jpg, .jpeg, and .png image formats for input. Data can be exported in JSON file.

- **Hasty.ai (https://hasty.ai/)**
  Hasty is an image annotation software written in Python. It is a simple image annotation tool that makes teaching machines easier, faster, and more accurate. Hasty is proprietary software that supports a data-centric machine learning platform. It also supports vector labeling, provides pre-generated labels, and label class attributes prediction. The software provides full security to the user, i.e., the data is not shared or used to train the model of any other user. It can export data in JSON, Pascal VOC and COCO format.

- **Hive (https://thehive.ai/)**
  Hive is enterprise image annotation software written in Java. It that provides cloud-based enterprise AI solutions. Hive is a project management proprietary software that allows the users to collaboratively annotate and markup documents in real time. It supports multi-frame object tracking, contouring, and 3D panoptic segmentation. Hive also provides hotkeys for quick access.

- **ImageTagger (https://github.com/bit-bots/imagetagger)**
  ImageTagger is an image annotation software written in Python and HTML. It is an open-source platform that provides collaborative online as well as offline image annotation. Imagetagger supports the collection of filtered images from multiple image sets and provides image metadata. It can be used to share labeled image data for supervised learning in object recognition. The software is easy to use and does not require any installation.

- **Label360 (https://github.com/ailabstwLabel360)**
  Label360 is a software that provides an annotation interface for labeling instance-aware semantic labels on panoramic full images. It is a freely available user-friendly online software that is not required to be installed. The software is used to solve the distortion and instance matching issues across different viewing aspects in spherical image annotations. A post-processing algorithm is introduced to generate distortion-free annotations on equirectangular images.

- **Labelbox (https://labelbox.com/)**
  Labelbox is an image annotation software written in JavaSript and Python. It is a user-friendly software that is available online. There is no need to install the software. Labelbox is engineered for precision and speed. It supports automated labeling processes and training models for active learning. The tool can import data from any source. The software runs on Linux, Windows, and Mac. The tool also provides hotkeys.

- **LabelImg (https://github.com/tzutalin/labelImg)**
  LabelImg is an image annotation software written in Python. It is a free, open-source tool for graphically labeling images. LabelImg is an easy and free way to label a few hundred images. The software is suitable only for object detection or object localization by solely creating rectangular boxes. LabelImg supports labeling in VOC, XML or YOLO text file format. The software runs on Linux, Windows, and Mac. The tool also provides hotkeys.

- **LabelMe (http://labelme.csail.mit.edu/Release3.0/)**
  LabelMe is an image annotation software written in Python. It is a freely available user-friendly tool. LabelMe is an online image polygonal annotation tool that is used to build image databases. Functions can be customized by users and the tool can be accessed from anywhere, i.e., there is no need to install it. LabelMe supports object detection, classification, and segmentation. The software runs on Linux, Windows, and Mac. The tool also provides hotkeys.

- **Label Studio (https://github.com/ heartexlabs/label-studio)**
  Label Studio is an image annotation software written in Python. It is a freely available online tool. The user interface is simple and there is no need to install it. Label Studio is used to associate labeling activities with a specific user and manage the data in a user interface. The software supports image classification, object detection, and semantic segmentation. The tool uses machine learning models to pre-label and optimizes the process. Label Studio is available as a community edition open-source data labeling tool. It is also available as a paid version with extended functionality and support. The software runs on Linux, Windows, and Mac.

- **MakeSense.ai (https://www.makesense.ai/)**
  Makesense.ai is an image annotation software written in TypeScript. It is a free-to-use online tool for labeling photos and does not require any complicated installation. The software is perfect for small computer vision deep learning projects, making the process of preparing a dataset much easier and faster. Prepared labels can be downloaded in one of the multiple supported formats. SSD model is pretrained on the COCO dataset, which enables drawing boxes on photos and suggesting labels. The PoseNet model is a visual model that can be used to estimate the pose of a person in an image by estimating where key body joints are. The software can run on any operating system.
- **Playment (https://www.playment.io/)**
  Playment is a no-code, self-serve data labeling platform that is used to create diverse and high-quality ground truth datasets. It is a web-based labeling platform that eliminates inefficiencies for the annotator and the project manager via machine learning based annotation tool. It provides easy-to-use workflow management software.

- **RectLabel (https://rectlabel.com/)**
  RectLabel is an image annotation software written in Python. It supports bounding box object detection and segmentation. It can read/write in Pascal VOC and YOLO text format and can export to a YOLO, COCO JSON and CSV formats. The software supports editing points to fit the shape and automatic labeling using core machine learning models. RectLabel provides settings for objects, attributes, hotkeys, and labeling fast. Object/attribute names and image names can be searched. The software can run on iPadOS and windows.

- **Scale AI (https://scale.com/)**
  Scale AI is an image annotation software used to accelerate the development of artificial intelligence by providing a data-centric, end-to-end solution to manage the entire machine learning lifecycle. The tool can be used to annotate large volumes of the 3D sensor, image, and video data at high throughput. Scale AI supports object detection and classification in images.

- **Scalabel (https://github.com/scalabel/)**
  Scalable is an image annotation software written in TypeScript and Python. It is an open-source web annotation tool that supports semi-automated annotation. The is no need to install the software. It supports both 2D and 3D data labeling. The tool is used for 2D bounding box detection, 2D polygon/polyline, 3D bounding box detection and tracking annotation on point clouds. The software also provides real-time session synchronization for seamless collaboration, and semi-automatic annotation with label pre-loading. Deep learning models can be used to assist annotations.

- **Segments.ai (https://segments.ai/)**
  Segments.ai is an image annotation software written in Jupyter Notebook and Python. It provides a platform for smart labeling to manage, label, and improve datasets. The software supports semantic, instance, and panoptic segmentation. Machine learning-powered labeling tools Superpixels and Autosegment are used for labeling the images.

- **SuperAnnotate (https://www.superannotate.com/)**
  SuperAnnotate is an image annotation software written in Python. It provides an end-to-end platform to annotate and manage ground truth data. It supports smart segmentation and annotates multiple data formats. It provides advanced and user-friendly tools for more accurate annotations. It runs on Windows, Linux, and MacOS.

- **Supervise.ly (https://supervise.ly/)**
  Supervise.ly is an image annotation software written in Python. It is an all-in-one labeling platform with a turnkey solution. It can be used to develop faster and better artificial intelligence with an on-premise and enterprise-grade ecosystem, and build custom AI solutions as per the user needs. The software provides easy and convenient access to state-of-the-art models and machine learning tools.

- **TaQadam (https://taqadam.io/image-annotation/)**
  TaQadam is an image annotation software written in Python and Java. It transforms imagery collected via satellite and drones into valuable insights. The tool can be used to make visual data AI-ready. It is the platform for image annotation and crowdsourced geospatial imagery analysis that brings agility to the data training pipeline.

- **V7 (https://www.v7labs.com/)**
  V7 is an image annotation software written in Python. It can be used by non-technical users. It supports medical image annotations and provides dataset management. The tool is used to create pixel-perfect annotations on any data format with class-agnostic, automated tools. It uses neural networks to create ground truth and delightful UX.

- **VOTT or Visual Object Tagging Tool (https://github.com/microsoft/VOTT)**
  VoTT is an image annotation software written in TypeScript. It is a free, open-source annotation tool. Visual Object Tagging Tool is an electron app for building end-to-end object detection models from images. The software can label images, it provides an extensible model for importing data from local or cloud storage providers, and an extensible model for exporting labeled data to local or cloud storage providers. VoTT helps facilitate an end-to-end machine learning pipeline. Supervised learning is provided to train computer vision models.

- **VGG Image annotator or VIA (https://www.robots.ox.ac.uk/~vgg/software/via/)**
  VIA is an image annotation software that is based solely on HTML and JavaScript. It is a simple and standalone manual annotation software. VIA runs in a web browser and does not require any installation or setup. VIA is an open-source project.

Capabilities of IATs
The types of image annotation tasks performed by each of the tools is illustrates in Table 1.

Table 1: Image annotation functions supported by the Image annotation Tools

S.No.	**Image Annotation Tools**	**Computer vision tasks that require Image Annotation**			
		Image Classification	**Object Detection and Recognition**	**Image Segmentation**	**Boundary Recognition**
1.	Anno-Mage	×	✓	×	×
4.	CVAT	✓	✓	✓	×
5.	Dataturks	×	✓	✓	×
6.	Diffgram	×	✓	✓	×
7.	Hasty.ai	✓	✓	✓	×
8.	Hive	×	✓	✓	×
9.	ImageTagger	×	×	×	×
10.	Label360	×	×	×	✓
11.	Labelbox	✓	✓	✓	×
12.	LabelImg	✓	✓	×	×
13.	LabelMe	✓	✓	✓	✓

S.No.	**Image Annotation Tools**	**Computer vision tasks that require Image Annotation**			
		Image Classification	**Object Detection and Recognition**	**Image Segmentation**	**Boundary Recognition**
14.	Label Studio	✓	✓	✓	×
15.	MakeSense	×	✓	×	×
16.	Playment	×	✓	✓	×
17.	Prodigy	✓	✓	×	×
18.	RectLabel	×	✓	✓	×
19.	Scale AI	✓	✓	✓	×
20.	Scalabel	×	×	✓	×
21.	Segments.ai	×	×	×	×
22.	SuperAnnotate	×	✓	✓	×
23.	Supervise.ly	✓	✓	✓	✓
24.	TaQadam	✓	×	✓	×
25.	V7	✓	✓	✓	×
26.	VOTT	×	✓	×	×
27.	VIA	×	✓	×	✓

Image annotation techniques supported by each of the selected tools are given in Table 2.

Table 2: Image annotation techniques supported by the Image annotation Tools

S.No.	**Image Annotation Tools**	**Image Annotation Techniques**					
		Bounding Box Anno	**Polygon/ Contour Anno**	**Line Anno**	**Landmark Anno**	**Semantic Seg Anno**	**Cubic Anno**
1.	Anno-Mage	✓	×	×	×	×	✓
4.	CVAT	✓	✓	×	✓	×	×
5.	Dataturks	×	✓	×	×	×	×
6.	Diffgram	✓	✓	×	×	×	×
7.	Hasty.ai	✓	✓	×	✓	×	×
8.	Hive	✓	✓	×	✓	×	×
9.	ImageTagger	✓	✓	×	✓	✓	×
10.	Label360	✓	✓	✓	✓	✓	✓
11.	Labelbox	✓	✓	×	✓	×	×
12.	LabelImg	✓	×	×	×	×	×
13.	LabelMe	✓	✓	✓	✓	✓	✓
14.	Label Studio	×	×	×	×	×	×
15.	MakeSense	✓	✓	×	✓	✓	×
16.	Playment	✓	✓	×	✓	✓	×
17.	Prodigy	✓	✓	×	×	×	×
18.	RectLabel	✓	✓	✓	✓	✓	✓
19.	Scale AI	✓	✓	×	✓	×	×
20.	Scalabel	✓	✓	×	✓	✓	✓
21.	Segments.ai	✓	×	×	✓	✓	×
22.	SuperAnnotate	✓	✓	×	✓		✓
23.	Supervise.ly	✓	✓	✓	✓	✓	✓
24.	TaQadam	✓	✓	×	✓	×	×
25.	V7	×	×	×	×	×	×
26.	VOTT	✓	×	×	×	×	×
27.	VIA	✓	✓	×	✓	×	×

Table 3 describes the supported formats of images and datasets, which are used by image annotation tools. Some tools can directly import images from datasets like Pascal VOC, COCO, and YOLO while some image annotation tools can import and export images in a certain image format like PNG, JPEG, or BMP. Exported annotation can also be accessed through CSV, XML, and ZIP file types.

Table 3: Formats supported by image annotation tools

SNo	Image annotation Tools	Import from			Export to				Can be accessed through		
		Pascal VOC	COCO	YOLO	BMP	TIFF	PNG	JPEG	XML	ZIP	CSV
1	Anno Mage	✓					✓				
4	CVAT										
5	Dataturks	✓	✓							✓	
6	Diffgram		✓								✓
7	Hasty.ai							✓			
8	Hive		✓								
9	ImageTagger			✓							
10	Label360*										
11	Labelbox		✓								✓
12	LabelImg	✓		✓							
13	LabelMe	✓						✓	✓		
14	Label Studio	✓	✓				✓		✓		✓
15	MakeSense	✓	✓	✓					✓		✓
16	Playment										✓
17	Prodigy						✓	✓			
18	RectLabel	✓	✓	✓							✓
19	Scale AI					✓	✓	✓			
20	Scalabel	✓	✓	✓							
21	Segments.ai		✓	✓			✓	✓			
22	SuperAnnotate				✓	✓	✓	✓			
23	Supervise.ly	✓					✓	✓	✓		
24	TaQadam		✓								
25	V7				✓	✓	✓	✓		✓	
26	VIA	✓	✓	✓					✓		✓
27	VOTT	✓	✓								✓

Table 4 illustrates the utilization of tools in various application areas by the researchers. This will enable the users in understanding the applicability of the tools to annotate the images.

Table 4: Utilization of tools

IAT	Utilization	In the field of
Hasty.ai	Bassiouny, Farhan, Maged & Awaad (2021) compared various computer vision algorithms to detect, localize and classify precious components in laptops to dismantle the laptops as non-destructively as possible. Hasty.ai was used to label the data with the bounding boxes method.	E-waste components detection
Hive	Yao et al. (2021) provided a dataset consisting of annotated photographs of non-human primates in naturalistic contexts obtained from the internet, National Primate Research Centers, and the Minnesota Zoo. The authors used Hive AI to annotate the landmarks from cropped images using bounding box.	Pose Tracking of Non-human Primates
Labelbox	Parraga-Alava et al. (2021) presented a dataset consisting of healthy and unhealthy (leaves with the presence of aphids and visible white spots) lemon leaf images. Each image included a set of annotations that identify the leaf, its health state, and the infestation severity according to the percentage of the affected area on it.	Plant disease recognition
	Šiaulys et al. (2021) presented annotated images of the bottom macrofauna obtained from underwater video. 2D mosaics were manually annotated by	Marine biology

IAT	Utilization	In the field of
	several experts using the Labelbox tool and co-annotations were refined using the SurveyJS platform.	
LabelImg	Zhu, Wu, Yang, Shen & Wu (2016) applied Faster R-CNN to train a detection network on the digital image database of books and implement automatic recognition and positioning of books. This provides great help for the study of the practical book retrieval system. Graphical image annotation tool LabelImg was used for annotating the location of books.	Education (Detection of books)
	Tabassum, Ullah, Al-Nur & Shatabda (2020) presented the 'Poribohon-BD' dataset for vehicle classification in Bangladesh. The dataset contains labeled and annotated images of Bangladeshi vehicles. LabelImg was used for annotating the images. This can be useful in accident avoidance, and traffic congestion prevention.	Traffic congestion prevention
	Xie et al. (2020) created a model for an autonomous car in traffic sign recognition. The collected dataset was annotated into different classes and labels with LabelImg. The annotations were used to crop out images and categorized them. This data was run through a deep learning algorithm called modified AlexNet.	Traffic Sign Recognition
	Menon, Abhishek, SN & George (2021) developed a real-time translation software for the conversion of hand gestures into natural languages LabelImg was used for annotating the images.	Sign language recognition
	Politikos, Fakiris, Davvetas, Klampanos & Papatheodorou (2021) presented an object detection approach to automatically detects seafloor marine litter using a Region-based Convolution Neural Network. The neural network is trained on imagery with manually annotated litter categories and then evaluated on an independent part of the dataset.	Underwater imaging
	Wang & Xiao (2021) employed deep convolutional neural network models for potato surface defect detection. To train the model, each image was manually labeled in the dataset. LabelImg was used to draw a bounding box around a potato object in an image and mark its category.	Potato Surface Defect Detection
Label Studio	Blattner, Mommert & Borth (2021) investigated the feasibility of estimating hourly commercial vehicle traffic rates from freely available Sentinel2 satellite imagery. A model was trained to detect individual commercial vehicles in satellite images and the outcome was used to predict hourly commercial vehicle traffic rates. All visible commercial vehicles were annotated using Label Studio on the freeway with rectangular bounding boxes. The created dataset consisted of labeled images with annotations on individual commercial vehicles.	Traffic Detection
LableMe	Marder et al. (2015) addressed the challenges of detecting and identifying multiple objects on store shelves for monitoring planogram compliance. The dataset comprised photos of 60 in-store shelves. In those shelf captures, appearances of different products were annotated using LabelMe.	Retail
MakeSense	Tavakkol et al. (2020) introduce the modular and scalable design of Kartta Labs, an open source, open data, and scalable system for virtually reconstructing cities from historical maps and photos to reconstruct a city from historical maps and photos. The historical photos were annotated using MakeSense to identify building facades and then to identify the structural details of each facade.	Reconstructing structures of historical buildings
Playment	Schmidt (2019) considered labeled images of autonomous vehicles in California which were stills taken from videos shot in traffic. These were then manually annotated using Playment. Bounding boxes and semantic segmentation maps were used for annotation. This will enable to predict how vehicles or people will behave in traffic.	Autonomous vehicles
Rectlabel	Jamtsho, Riyamongkol & Waranusast (2020) presented the real-time Bhutanese license plate localization using YOLO. Before training the YOLO model, the bounding box was generated for each class. Two classes (Vehicle and Plate) were annotated per image using the RectLabel.	License plate localization
	Wei et al. (2020) evaluated the performance and generalizability of a deep neural network for colorectal polyp classification on histopathologic slide	Cancer Diagnosis

IAT	Utilization	In the field of
	images using a multi-institutional data set. The images were annotated by gastrointestinal pathologists. For 157 whole-slide images in the training set, 2 of the gastrointestinal pathologists identified the polyps on the slides and used the Rectlabel annotation tool to manually annotate rectangular bounding boxes around polyps and normal tissue as regions of interest for model training.	
	Meda, Milla & Rostad (2021) demonstrated object detection in identifying rickets and normal wrists on pediatric wrist radiographs. Two pediatric radiologists reviewed the images and categorized them as either rickets or normal. Images were annotated by drawing a labeled bounding box around the distal radial and ulnar metaphases. RectLabel was used for annotation. Annotation data was stored in XML files and were exported to a CSV file.	Rickets Diagnosis
SuperAnnotate	Panferov, Tailakov & Donets (2020) created a dataset comprised of images of segmented core specifically for recognizing Rocks Lithology. SuperAnnotate tool was used to label the images conveniently.	Biodiversity
	Barragan, Chanci, Yu & Wachs (2021) developed a semi-autonomous robotic suction assistant, which was integrated with a Da Vinci Research Kit. An autonomous control based on a deep learning model was used to segment and identify the location of blood accumulations. The images were annotated using SuperAnnotate.	Emergency teleoperated suction
Supervise.ly	Iglovikov, Rakhlin, Kalinin & Shvets (2018) described an automated deep learning approach for bone age assessment. Masks were manually labeled using Supervisely and were used to train the U-Net model.	Bone age assessment
	Ohee & Asif (2020) developed a method to detect tigers in real-time visually using the YOLOv3 algorithm. A dataset of tiger images was collected containing tigers with different positions and orientations. YOLOv3 has been trained and validated using Supervise.ly image annotation and clustered to evaluate the real-time tiger detection system.	Tiger Detection
Taqadam	Persson (2019) analyzed the drone imagery of flooded regions using deep neural networks to facilitate disaster prevention and response. Deep neural networks have been used to do image segmentation of buildings, roads, and water.	Flooded Regions Detection
VoTT	Naseer, Baro, Khan & Gordillo (2020) applied Faster RCNN models for object detection and classification. The images were labeled manually using Microsoft VOTT to record the ground truth annotations and was saved in Pascal VOC format. The saved XML annotation file contains the image name, class name (Nephrops), and bounding box details of each object in the image.	Lobster Nephrops norvegicus stock
	Ramesh et al. (2021) discussed the use of Microsoft VoTT for comprehensive and customized data labeling. Customized annotations of the optic nerve head and retinal nerve fiber layer images enabled in identifying glaucomatous discs and in predicting segmentation of the glaucomatous cup in the background fundus separately.	Glaucoma diagnosis
	Sezen, Turhan & Sengul (2021) developed a framework to generate natural language descriptions of images within a controlled environment. The developed semantic image annotation framework combined deep learning models and aligned annotation results derived from the instances of the ontology classes to generate sentential descriptions of images.	Sports
VIA	Undit, Hassan & Zin (2021) reconstructed and trained the Mask Region Convolutional Neural Networks with the collection of annotated unmarked road image datasets. The model was trained to detect the unmarked road with various road conditions and perform semantic segmentation. The unmarked road with various background conditions images were annotated using VIA. The model enabled to generate the lane departure warning.	Lane Departure Warning System

Table 5 displays the comparison of the tools.

Table 5: Comparison Chart for image annotation tools

IAT	Anno Mage	CVAT	Dataturks	Diffgram	Hasty.ai	Hive	ImageTagger	Label360	Labelbox	LabelImg	LabelMe	Label Studio	MakeSense	Playment	Prodigy	RectLabel	Scale AI	Scalabel	Segments.ai	SuperAnnotate	Supervise.ly	TaQadam	V7	VOTT	VIA
UI	UF	C	C	UF	UF	UF	UF	UF	UF	UF	UF	UF	UF	C	C	C	UF	UF	UF	UF	UF	UF	UF	C	
Installa	Easy	Easy	NR	Easy	NR	C	NR	NR	NR	C	C	NR	NR	NR	Easy	C	NR	NR	NR	Easy	NR	Easy	NR	C	Easy
Hotkey						Y			Y	Y	Y	Y	Y			Y				Y				Y	Y
CP	Y	Y	Y	Y					Y	Y	Y	Y	Y			N					Y			Y	Y
Mode		O	B	B			O		O	OFF	O	O	O			OFF		O			O			B	B
Availability	F	F	F	F	P	P	F	F	P	F	F	F	F	P	P	F	P	F	P	P	F	P	P	F	P
Data types	I	I, V	I, V, T	I, V, T	I	I, V, T	I	I	I, V, T, G	I	I	I, T, A	I	I, V, S	I	I	I, V, S	I	I	I	I, V	I, G	I, V	I, V	I, V, A

Abbreviation used in the table

UF- User friendly	O- Online	V-Video
C- Complex	OFF-Offline	A-Audio
NR- Not required	S- Sensing	I- Image
Y- Yes	P- Proprietary	G- Geospatial
F- Free	B- Both	

The criteria to choose an appropriate image annotation tools are:

- Efficiency- In terms of user-friendliness, documentation, ease of installation and hot-key support to save time and improve annotation quality.
- Functionality- In terms of supporting cross-platform and the required annotation function
- Formatting- In terms of import image format, dataset support format or the export format supported by the image annotation tools.
- Application-It means that the user is able to run the annotation tool in online or offline mode as preferred by him.
- Price-Usually, a user may prefer the tool that is freely available. Table 4 depicts the freely available tools.

CONCLUSION

The paper presents the critical analysis of the selected image annotation tools which are used by preceding researchers in different computer vision tasks like labeling, annotating, object tracking and segmentation. Due to magnificence of image datasets, the development of image annotation tools will probably keep growing in coming decades because of exponential increase in digital images. Therefore, a few suggested tools provide promising results and few might not. The implementation of the tools are limited in the area of medical, security, transportation and forensics. In current scenario, many image annotation tools uses the boundary recognition technique as shown in Table 1, whereas in future trends, it can be used to enhance the decision making using pixel wise annotation techniques which provides more accuracy in terms of annotating the object correctly. Still a lot of blank space is available for future development and hence, new tools are being developed tremendously.

References

[1] Andrade-Hoz, J., Vega-Gorgojo, G., Ruano, I., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Bravo, F., & Ordóñez, C. (2021, September). Educawood: A socio-semantic annotation system for environmental education. In *European Conference on Technology Enhanced Learning* (pp. 368-372). Springer, Cham.

[2] Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., & Sekimoto, Y. (2021). RDD2020: An annotated image dataset for automatic road damage detection using deep learning. *Data in brief*, *36*, 107133.

[3] Barragan, J. A., Chanci, D., Yu, D., & Wachs, J. P. (2021, August). SACHETS: Semi-autonomous cognitive hybrid emergency teleoperated suction. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)* (pp. 1243-1248). IEEE.

[4] Bassiouny, A. M., Farhan, A. S., Maged, S. A., & Awaad, M. I. (2021, May). Comparison of Different Computer Vision Approaches for E-waste Components Detection to Automate E-waste Disassembly. In *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)* (pp. 17-23). IEEE.

[5] Blattner, M., Mommert, M., & Borth, D. (2021). Commercial Vehicle Traffic Detection from Satellite Imagery with Deep Learning. Tackling Climate Change with Machine Learning Workshop at ICML 2021.

[6] Buendía, F., Gayoso-Cabada, J., & Sierra, J. L. (2020, October). An Annotation Approach for Radiology Reports Linking Clinical Text and Medical Images with Instructional Purposes. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 510-517).

[7] Chang, L., Zhuang, W., Wu, R., Feng, S., Liu, H., Yu, J., ... & Zhang, J. (2020). DARWIN: A Highly Flexible Platform for Imaging Research in Radiology. *arXiv preprint arXiv:2009.00908*.

[8] Chiatti, A., Motta, E., Daga, E., & Bardaro, G. (2020). Fit to measure: Reasoning about sizes for robust object recognition. *arXiv preprint arXiv:2010.14296*.

[9] Chiu, M. T., Xu, X., Wei, Y., Huang, Z., Schwing, A. G., Brunner, R., ... & Shi, H. (2020). Agriculture-vision: A large aerial image database for agricultural pattern analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2828-2838).

[10] Dassi, R., Kamal, S., & Babu, B. S. (2021, December). E-Waste Detection and Collection assistance system using YOLOv5. In *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)* (pp. 1-6). IEEE.

[11] Ekanayake, P., Deng, Z., Yang, C., Hong, X., & Yang, J. (2019, July). Naïve approach for bounding box annotation and object detection towards smart retail systems. In *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage* (pp. 218-227). Springer, Cham.

[12] Fiedler, N., Bestmann, M., & Hendrich, N. (2018, June). Imagetagger: An open source online platform for collaborative image labeling. In *Robot World Cup* (pp. 162-169). Springer, Cham.

[13] Ha, Q., Watanabe, K., Karasawa, T., Ushiku, Y., & Harada, T. (2017, September). MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5108-5115). IEEE.

[14] Haug, S., & Ostermann, J. (2014, September). A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In *European conference on computer vision* (pp. 105-116). Springer, Cham.

[15] Iglovikov, V. I., Rakhlin, A., Kalinin, A. A., & Shvets, A. A. (2018). Paediatric bone age assessment using deep convolutional neural networks. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 300-308). Springer, Cham.

[16] Jamtsho, Y., Riyamongkol, P., & Waranusast, R. (2020). Real-time Bhutanese license plate localization using YOLO. *ICT Express*, *6*(2), 121-124.

[17] Jin, S., Xu, L., Xu, J., Wang, C., Liu, W., Qian, C., ... & Luo, P. (2020, August). Whole-body human pose estimation in the wild. In *European Conference on Computer Vision* (pp. 196-214). Springer, Cham.

[18] Jin, Y., Li, J., Ma, D., Guo, X., & Yu, H. (2017, July). A semi-automatic annotation technology for traffic scene image labeling based on deep learning preprocessing. In *2017 IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC)* (Vol. 1, pp. 315-320). IEEE.

[19] Khan, N., Ullah, A., Haq, I. U., Menon, V. G., & Baik, S. W. (2021). SD-Net: Understanding overcrowded scenes in real-time via an efficient dilated convolutional neural network. *Journal of Real-Time Image Processing*, *18*(5), 1729-1743.

[20] Kim, J., & Park, C. (2017). End-to-end ego lane estimation based on sequential transfer learning for self-driving cars. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 30-38).

[21] Lee, J., Nazki, H., Baek, J., Hong, Y., & Lee, M. (2020). Artificial intelligence approach for tomato detection and mass estimation in precision agriculture. *Sustainability*, *12*(21), 9138.

[22] Li, R., Ranipeta, A., Wilshire, J., Malczyk, J., Duong, M., Guralnick, R., ... & Jetz, W. (2021). A cloud-based toolbox for the versatile environmental annotation of biodiversity data. *PLoS biology*, *19*(11), e3001460.

[23] Liang-Han, L. (2020). Label360: An annotation interface for labelling instance-aware semantic labels on panoramic full images. DOI： 10.6342/NTU202001310

[24] Marder, M., Harary, S., Ribak, A., Tzur, Y., Alpert, S., & Tzadok, A. (2015). Using image analytics to monitor retail store shelves. *IBM Journal of Research and Development*, *59*(2/3), 3-1. (LableMe)
[25] Meda, K. C., Milla, S. S., & Rostad, B. S. (2021). Artificial intelligence research within reach: an object detection model to identify rickets on pediatric wrist radiographs. *Pediatric Radiology*, *51*(5), 782-791.
[26] Menon, A. G., Abhishek, A., SN, A., & George, A. (2021). SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS IN MACHINE LEARNING. *International Journal of Advanced Research in Computer Science*, *12*.
[27] Morya, A. K., Gowdar, J., Kaushal, A., Makwana, N., Biswas, S., Raj, P., ... & Salodia, U. P. (2021). Evaluating the Viability of a Smartphone-Based Annotation Tool for Faster and Accurate Image Labelling for Artificial Intelligence in Diabetic Retinopathy. *Clinical Ophthalmology (Auckland, NZ)*, *15*, 1023.
[28] Nagori, M., Jachak, R. S., & Chaudhari, P. P. (2019, February). A framework for segregating solid waste by employing the technique of image annotation. In *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)* (pp. 1-6). IEEE.
[29] Naseer, A., Baro, E. N., Khan, S. D., & Gordillo, Y. V. (2020, October). Automatic Detection of Nephrops norvegicus Burrows in Underwater Images Using Deep Learning. In *2020 Global Conference on Wireless and Optical Technologies (GCWOT)* (pp. 1-6). IEEE.
[30] Nguyen, C., Huynh, M. T., Tran, M. Q., Nguyen, N. H., Jain, M., Vo, T. D., ... & Truong, S. Q. H. (2021, August). GOAL: Gist-set Online Active Learning for Efficient Chest X-ray Image Annotation. In *Medical Imaging with Deep Learning* (pp. 545-553). PMLR.
[31] Nogueira, V., Oliveira, H., Silva, J. A., Vieira, T., & Oliveira, K. (2019, October). RetailNet: A deep learning approach for people counting and hot spots detection in retail stores. In *2019 32nd SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)* (pp. 155-162). IEEE.
[32] Ohee, M. N. S., & Asif, M. A. G. (2020). Real-Time Tiger Detection u sing YOLOv3. *International Journal of Computer Applications*, *975*, 8887.
[33] Panferov, V., Tailakov, D., & Donets, A. (2020, November). Recognition of Rocks Lithology on the Images of Core Samples. In *2020 Science and Artificial Intelligence conference (SAI ence)* (pp. 54-57). IEEE.
[34] Parraga-Alava, J., Alcivar-Cevallos, R., Morales Carrillo, J., Castro, M., Avellán, S., Loor, A., & Mendoza, F. (2021). LeLePhid: An Image Dataset for Aphid Detection and Infestation Severity on Lemon Leaves. *Data*, *6*(5), 51.
[35] Pazzaglia, G., Mameli, M., Rossi, L., Paolanti, M., Mancini, A., Zingaretti, P., & Frontoni, E. (2021, January). People counting on low cost embedded hardware during the SARS-CoV-2 pandemic. In *International Conference on Pattern Recognition* (pp. 521-533). Springer, Cham.
[36] Pereira, J. W., & Ribeiro, M. X. (2021, June). Semantic annotation and classification of mammography images using ontologies. In *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 378-383). IEEE.
[37] Persson, T. (2019). Analyzing Drone Imagery of Flooded Regions with Deep Neural Networks. https://www.diva-portal.org/smash/get/diva2:1376296/
[38] Philbrick, K. A., Weston, A. D., Akkus, Z., Kline, T. L., Korfiatis, P., Sakinis, T., ... & Erickson, B. J. (2019). RIL-contour: a medical imaging dataset annotation tool for and with deep learning. *Journal of digital imaging*, *32*(4), 571-581.
[39] Politikos, D. V., Fakiris, E., Davvetas, A., Klampanos, I. A., & Papatheodorou, G. (2021). Automatic detection of seafloor marine litter using towed camera images and deep learning. *Marine Pollution Bulletin*, *164*, 111974.
[40] Ramesh, P. V., Ramesh, S. V., Aji, K., Ray, P., Tamilselvan, S., Parthasarathi, S., ... & Rajasekaran, R. (2021). Modeling and mitigating human annotations to design processing systems with human-in-the-loop machine learning for glaucomatous defects: The future in artificial intelligence. *Indian Journal of Ophthalmology*, *69*(10), 2892.
[41] Rojas, C., Rodríguez-Sánchez, A., & Renaudo, E. (2022). Deep Learning for Fast Segmentation of E-waste Devices' Inner Parts in a Recycling Scenario. In *International Conference on Pattern Recognition and Artificial Intelligence* (pp. 161-172). Springer, Cham.
[42] Santos, T. T., de Souza, L. L., dos Santos, A. A., & Avila, S. (2020). Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Computers and Electronics in Agriculture*, *170*, 105247.
[43] Schmidt, F. A. (2019). *Crowdsourced production of AI training data: how human workers teach self-driving cars how to see* (No. 155). Working Paper Forschungsförderung.
[44] Schmidtke, L., Vlontzos, A., Ellershaw, S., Lukens, A., Arichi, T., & Kainz, B. (2021). Unsupervised human pose estimation through transforming shape templates. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2484-2494).
[45] Sezen, A., Turhan, C., & Sengul, G. (2021). A Hybrid Approach for Semantic Image Annotation. *IEEE Access*, *9*, 131977-131994.
[46] Shao, J., Li, J., Kong, W., Liu, S., Wu, J., & Wu, H. (2021, June). A crowdsourcing framework for retinal image semantic annotation and report documentation with deep learning enhancement. In *Journal of Physics: Conference Series* (Vol. 1955, No. 1, p. 012037). IOP Publishing.

[47] Šiaulys, A., Vaičiukynas, E., Medelytė, S., Olenin, S., Šaškov, A., Buškus, K., & Verikas, A. (2021). A fully-annotated imagery dataset of sublittoral benthic species in Svalbard, Arctic. *Data in brief*, *35*, 106823.
[48] Sie, E., & Vasisht, D. (2022, May). RF-Annotate: Automatic RF-Supervised Image Annotation of Common Objects in Context. In *2022 International Conference on Robotics and Automation (ICRA)* (pp. 2590-2596). IEEE.
[49] Sudars, K., Jasko, J., Namatevs, I., Ozola, L., & Badaukis, N. (2020). Dataset of annotated food crops and weed images for robotic computer vision control. *Data in brief*, *31*, 105833.
[50] Tabassum, S., Ullah, S., Al-Nur, N. H., & Shatabda, S. (2020). Poribohon-BD: Bangladeshi local vehicle image dataset with annotation for classification. *Data in brief*, *33*, 106465.
[51] Tan, W., Duan, Q., Yao, L., & Li, J. (2022). A sensor combination based automatic sorting system for waste washing machine parts. *Resources, Conservation and Recycling*, *181*, 106270.
[52] Tavakkol, S., Han, F., Mayer, B., Phillips, M., Shahabi, C., Chiang, Y. Y., & Kiveris, R. (2020). Kartta labs: Collaborative time travel. *arXiv preprint arXiv:2010.06536*.
[53] Thumm, J., Masino, J., Knoche, M., Gauterin, F., & Reischl, M. (2020). A novel approach to label road defects in video data: semi-automated video analysis. *International Journal on Smart Sensing & Intelligent Systems*, *13*(1).
[54] Torres, M., & Qiu, G. (2013, September). Habitat classification using random forest based image annotation. In *2013 IEEE International Conference on Image Processing* (pp. 1491-1495). IEEE.
[55] Undit, H. J. A., Hassan, M. F. A., & Zin, Z. M. (2021, September). Vision-Based Unmarked Road Detection with Semantic Segmentation using Mask R-CNN for Lane Departure Warning System. In *2021 4th International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR)* (pp. 1-6). IEEE.
[56] Vargas Munoz, J. E., Tuia, D., & Falcão, A. X. (2021). Deploying machine learning to assist digital humanitarians: making image annotation in OpenStreetMap more efficient. *International Journal of Geographical Information Science*, *35*(9), 1725-1745.
[57] Wang, C., & Xiao, Z. (2021). Potato Surface Defect Detection Based on Deep Transfer Learning. *Agriculture*, *11*(9), 863.
[58] Wei, J. W., Suriawinata, A. A., Vaickus, L. J., Ren, B., Liu, X., Lisovsky, M., ... & Hassanpour, S. (2020). Evaluation of a deep neural network for automated classification of colorectal polyps on histopathologic slides. *JAMA network open*, *3*(4), e203398-e203398.
[59] Xie, D., Nuakoh, E., Chatterjee, P., Ghattan, A., Edoh, K., & Roy, K. (2020, February). Traffic Sign Recognition for Self-driving Cars with Deep Learning. In *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 207-217). Springer, Singapore.
[60] Yao, Y., Mohan, A. A., Bliss-Moreau, E., Coleman, K., Freeman, S. M., Machado, C. J., ... & Park, H. S. (2021). OpenMonkeyChallenge: Dataset and Benchmark Challenges for Pose Tracking of Non-human Primates. *bioRxiv*.
[61] You, J. A., Hu, Z., Peng, C., & Wang, Z. (2021). Generation and Annotation of Simulation-Real Ship Images for Convolutional Neural Networks Training and Testing. *Applied Sciences*, *11*(13), 5931.
[62] Zhang, D., Islam, M. M., & Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, *45*(1), 346-362.
[63] Zheng, Y. Y., Kong, J. L., Jin, X. B., Wang, X. Y., Su, T. L., & Zuo, M. (2019). CropDeep: the crop vision dataset for deep-learning-based classification and detection in precision agriculture. *Sensors*, *19*(5), 1058.
[64] Zhu, B., Wu, X., Yang, L., Shen, Y., & Wu, L. (2016, July). Automatic detection of books based on Faster R-CNN. In *2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC)* (pp. 8-12). IEEE.
[65] Zurowietz, M., & Nattkemper, T. W. (2020). Unsupervised knowledge transfer for object detection in marine environmental monitoring and exploration. *IEEE Access*, *8*, 143558-143568.
[66] Zurowietz, M., Langenkämper, D., Hosking, B., Ruhl, H. A., & Nattkemper, T. W. (2018). MAIA—A machine learning assisted image annotation method for environmental monitoring and exploration. *PloS one*, *13*(11), e0207498.