Data Labelling
&
Data Annotation

Technical Insights

C-LIVE Solutions Pvt. Ltd.
Website: www.c-live.in

DATA ANNOTATION AND LABELING

Annotation in machine learning is the process of labeling data, which could be in the form of text, images, audio, etc. Typically, annotation is done by a human.

Supervised machine learning algorithms learn from labeled data, and those data that has been tagged with labels. Programmers do not explicitly program machine learning algorithms on how to make decisions, and they program the models that learn from labeled data.

Data labeling, also called data annotation/tagging/classification, is the process of preparing tagged (i.e. labeled) data sets for machine learning. Machine learning models learn to recognize repetitive patterns in labeled data. After a sufficient amount of labeled data is processed, machine learning models can identify the same patterns in data that have not been labeled.

DIFFERENCE BETWEEN DATA ANNOTATION AND LABELING

In annotation the entire data are only annotated to make it important, while in labeling the metadata is also added to make it more descriptive to machine for natural language processing.

LIDAR AND RADAR DATA ANNOTATION

Lidar

High Resolution Radar
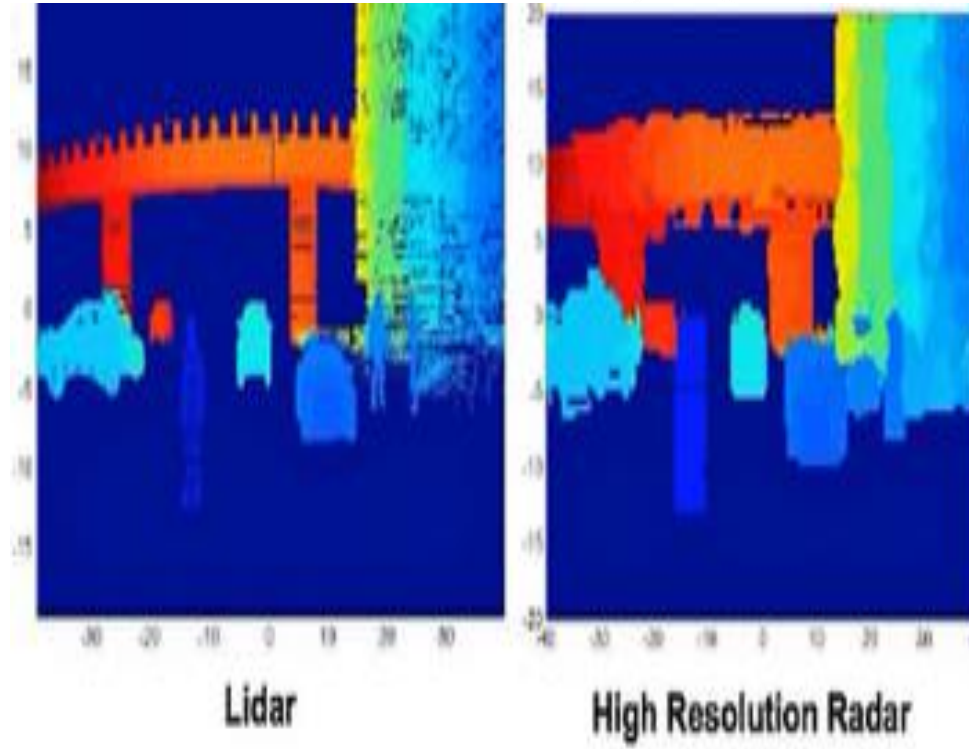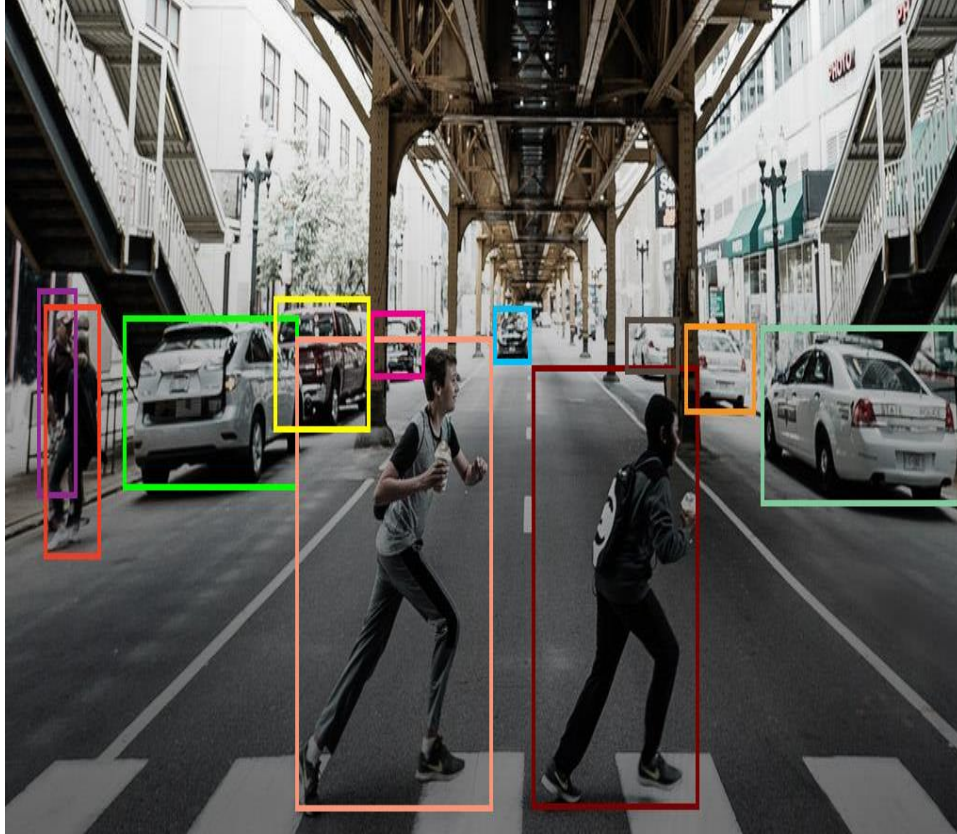
LIDAR/RADAR Annotation :Identifies objects in a 3D point cloud and draws bounding cuboids around the specified objects, returning the positions and sizes of these boxes

Point Annotation: Identifies the location of objects and draws points at specified locations, returning the locations of these points.

OBJECT DETECTION

Object detection is a computer vision technique that allows us to identify and locate objects in an image or video. With this kind of identification and localization, object detection can be used to count objects in a scene and determine and track their precise locations, all while accurately labeling them.
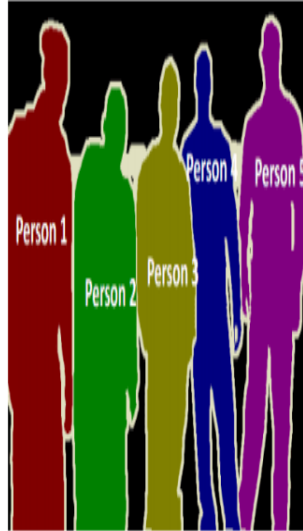
SEGMENTATION: SEMANTIC AND INSTANCE

Object Detection

Semantic Segmentation

Instance Segmentation

Semantic Segmentation : Is a technique that detects , for each pixel , the object category it belongs to , all object categories (labels) must be known to the model.

Instance Segmentation : Same as Semantic Segmentation, but dives a bit deeper, and it identifies , for each pixel, the object instance it belongs to. The main difference is that differentiates two objects with the same labels in comparison to semantic segmentation.

MEDICAL IMAGE TRANSCRIPTION

Medical image annotation is the process of labeling the medical imaging data like Ultrasound, MRI, and CT Scan, etc. for machine learning training. Apart from these radiologist images, other medical records available in the text formats are also annotated to make it understandable to machines through deep learning algorithms for accurate predictions.

IMAGE LABELLING

Image Labeling is the process of recognizing different entities in an image. One can recognize various entities like animals, plants, food, activities, colors, things, fictional characters, drinks etc. with Image Labeling.

Image labeling involves identifying an image as a whole, but it can also involve identifying various aspects within an image. For example, the process is straightforward for pictures that contain a single image, like a portrait photograph. Labeling an image can be more difficult for shots that include more details, like a wide-angle picture taken at a public place.

VIDEO LABELLING

Video annotation involves adding metadata to unlabeled video in order to train a machine learning algorithm. This metadata, also referred to as tags or labels, could be anything from a bounding box around a certain part of the image to full segmentation, where every pixel is annotated with its semantic meaning.

TEXT DECODING

Image with text

OCR → T

Scanned or Printed or handwritten document

Decoding is the process of translating print into speech by rapidly matching a letter or combination of letters (graphemes) to their sounds (phonemes) and recognizing the patterns that make syllables and words.

NEW YORK NEWS
Articles for you
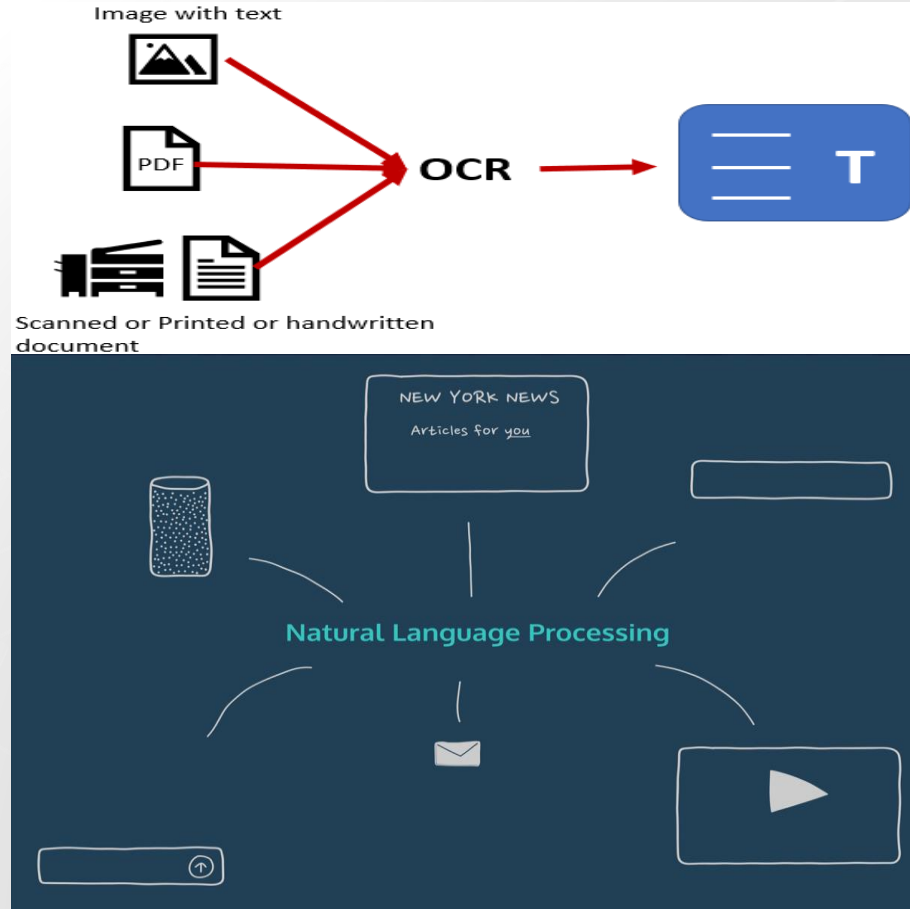
Natural Language Processing

IMAGE RECOGNITION

Image recognition is a computer vision task that works to identify and categorize various elements of images and/or videos. Image recognition models are trained to take an image as input and output one or more labels describing the image. The set of possible output labels are referred to as **target classes**. Along with a predicted class, image recognition models may also output a confidence score related to how certain the model is that an image belongs to a class.

For instance, if one want to build an image recognition model that automatically determined whether or not a dog was in a given image, the pipeline would, broadly speaking, look like this:

- Image recognition model trained on images that have been labeled as "dog" or "not dog"
- Model input: Image or video frame
- Model output: Class name (i.e. dog) with a confidence score that indicates the likelihood of that image containing that class of object.
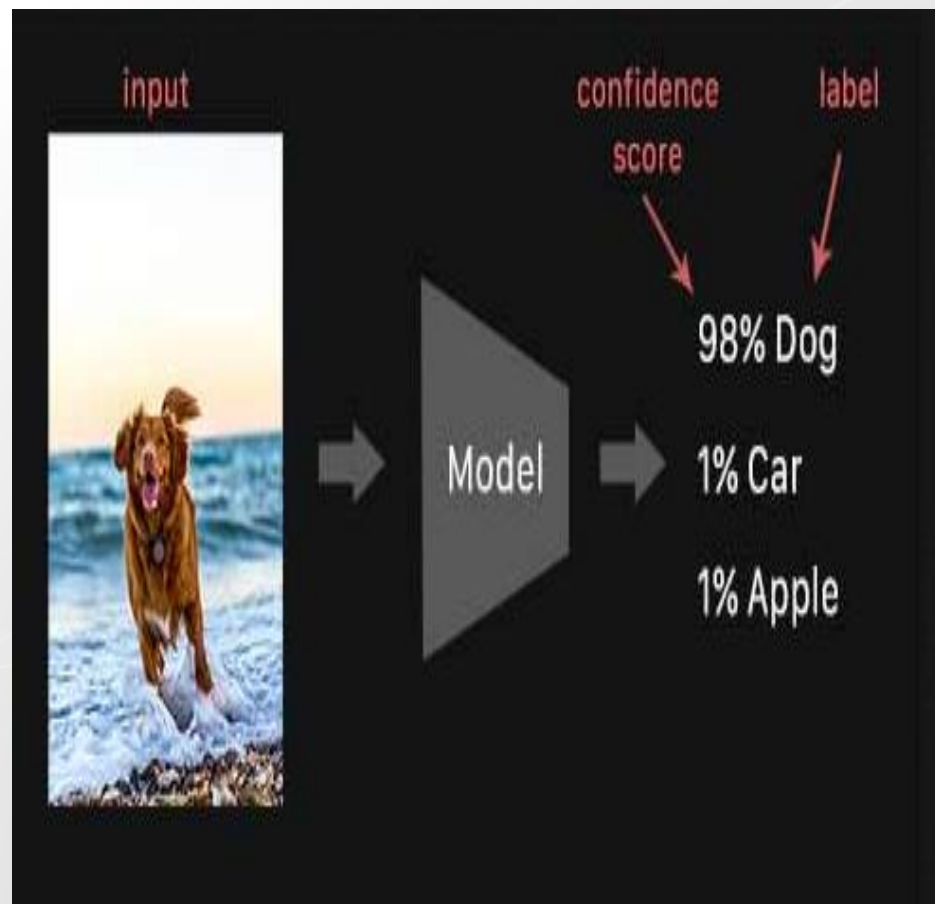
CLASSIFICATION AND CLUSTERING

Both Classification and Clustering are used for the categorization of objects into one or more classes based on the features. They appear to be a similar process as the basic difference is minute. In the case of Classification, there are predefined labels assigned to each input instances according to their properties whereas in clustering those labels are missing.

PARAMETER	CLASSIFICATION	CLUSTERING
Type	used for supervised learning	used for unsupervised learning
Basic	process of classifying the input instances based on their corresponding class labels	grouping the instances based on their similarity without the help of class labels
Need	it has labels so there is need of training and testing dataset for verifying the model created	there is no need of training and testing dataset
Complexity	more complex as compared to clustering	less complex as compared to classification
Example Algorithms	Logistic regression, Naive Bayes classifier, Support vector machines etc.	k-means clustering algorithm, Fuzzy c-means clustering algorithm, Gaussian (EM) clustering algorithm etc.

KEY CONTACTS:

Sanjib Dutta (Business Head)
@: sanjib@dtf-v.com

Nirbhay Kumar Singh (Project Head)
@: nirbhay@dtf-v.com

Sharmistha Dutt (VP-Commercial & Admin)
@: sd@dtf-v.com

Reach us:

Corporate Office:

LG 006, Springhouse Co-working, DLF Grand Mall, Near Sikanderpur Metro Station, Sector 28, M.G.Road, Gurugram, Haryana-122001

Regional Development Center:

Module 11, 12 & 13, Webel IT Tech Park, Phase 2, Near City Center, Durgapur, District - Burdwan, West Bengal - 713208

Module F, Webel IT Tech Park, Phase-1, Paribahan Nagar, Matigara, Siliguri, District-Darjeeling, West Bengal -734010

resources-icons-and-maps