**DAY 07**

**26.06.2023**

**Image recognition vs Image Classification**

|  |  |  |
| --- | --- | --- |
|  | Image Classification | Image Recognition |
| Definition | Assigning a label or category to an image | Comprehensive understanding of the visual content within an image |
| Complexity | Relatively simpler task | More complex and advanced task |
| Techniques | Feature extraction, machine learning models (e.g., CNNs) | Integration of multiple computer vision techniques (e.g., object detection) |
| Output | Single label or category | Detailed information about objects, attributes, and context |
| Example | Classifying images of animals into categories like "cat," "dog," or "bird" | Recognizing objects, detecting faces, understanding scenes |

**Annotations:**

Annotations play a crucial role in training and evaluating Convolutional Neural Networks (CNNs) for image-related tasks. Here are some common applications and uses of annotations in image CNNs:

* Image classification
* Object Detection
* Sematic and Instant Segmentation
* Facial Recognition
* Pose Estimation
* Image Captioning

**YOLO: Real – Time Object Detection**

You Only Look Once (YOLO) is a state-of-the-art, real-time object detection algorithm introduced in 2015 by[Joseph Redmon](https://arxiv.org/search/cs?searchtype=author&query=Redmon%2C+J),[Santosh Divvala](https://arxiv.org/search/cs?searchtype=author&query=Divvala%2C+S),[Ross Girshick](https://arxiv.org/search/cs?searchtype=author&query=Girshick%2C+R), and[Ali Farhadi](https://arxiv.org/search/cs?searchtype=author&query=Farhadi%2C+A) in their famous research paper “[**You Only Look Once: Unified, Real-Time Object Detection**](https://arxiv.org/abs/1506.02640)”.

***YOLO Architecture***

YOLO architecture is similar to[**GoogleNet**](https://arxiv.org/pdf/1409.4842.pdf)**.** It has overall 24 convolutional layers, four max-pooling layers, and two fully connected layers.

The architecture works as follows:

* Resizes the input image into 448x448 before going through the convolutional network.
* A 1x1 convolution is first applied to reduce the number of channels, which is then followed by a 3x3 convolution to generate a cuboidal output.
* The activation function under the hood is ReLU, except for the final layer, which uses a linear activation function.
* Some additional techniques, such as batch normalization and dropout, respectively regularize the model and prevent it from overfitting.

***Reference***

https://pjreddie.com/yolo/

<https://www.datacamp.com/blog/yolo-object-detection-explained/>

<https://www.altexsoft.com/blog/image-recognition-neural-networks-use-cases/>

<https://machinelearningmastery.com/object-recognition-with-deep-learning/>