CV / VLM

Unit 2: Introduction to Object Detection (OD)

2.2.1

Region-Based Object Detection

Sliding window approach

TIL-AI
TODAY I LEARNED AI

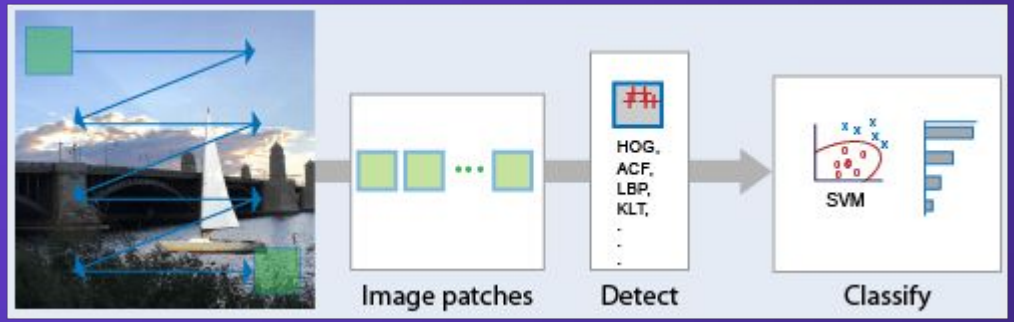
Object Detection Process:
Sliding Window Algorithm

Definition

- As per its name, sliding window involves the ‘sliding’ of a window/kernel across the image at every potential position (often, with overlaps).

Spatial Scan

- At every position, the model scans and analyzes the “image patches” for objects/patterns that it was trained to recognize, regardless of whether the window contains any relevant features.



(left) Sliding window algorithm process flow for Object Detection

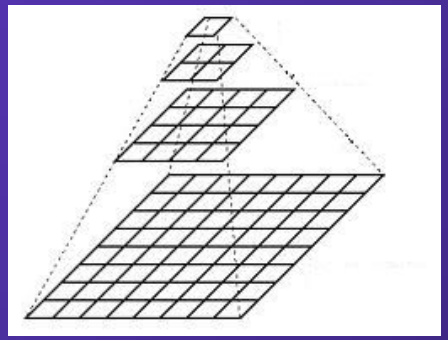
Anchor Boxes for Object Detection - MATLAB & Simulink (mathworks.com)

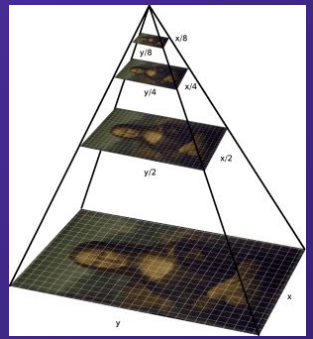
Object Detection Process:
Image Pyramid

- **Definition:**
  An image pyramid is a collection of images - that are successively downsampled until some desired stopping point is reached.

- **Functions:**
  “Image pyramid” (multi-scale representation of an image) to allow identification of objects at different scales of an image.
  - Each pyramid level can be used to detect objects at a different scale.

Different types of objects (swimming pool, tree) or even same type of object (buildings, human) can be of varying sizes (both due to physical properties and also distance to camera)





(left) An example of image pyramid, where each layers of image is downsized and optionally smoothed.

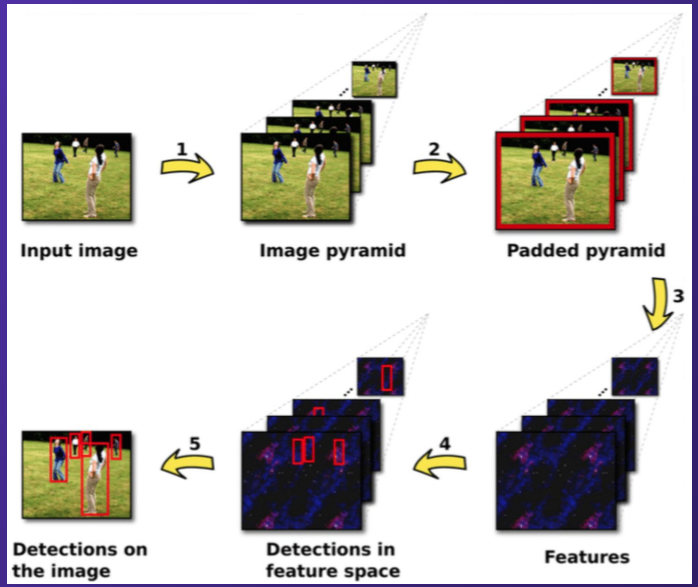
Anchor Boxes for Object Detection - MATLAB & Simulink (mathworks.com)

Object Detection Process:
Sliding Window + Image Pyramid

Flow

1. **Pyramid Creation**: Feed an input image into a pyramid creation process, resulting in a multi-resolution image representation.
2. **Padding**: Pad the pyramid of images to ensure consistent feature extraction.
3. **Feature Extraction**: Employ a sliding window technique to compute features at every position within the image pyramid.
4. **Object Detection**: Detect all objects based on the extracted features.
5. **Mapping**: Map the findings back to the original image, enabling object localization and classification.

By combining these techniques, the system can effectively detect objects in images, even when they vary in size or location.



Adapted from [Monitoring Human State in a Robotic Assistive Platform: Data Acquisition and Person Detection Systems]

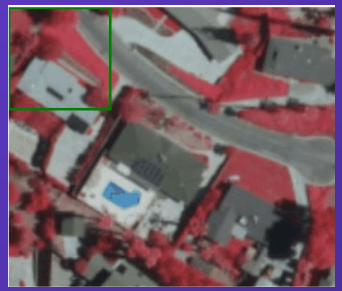
Limitations of Sliding Window

Object Size

- The minimum window size required to consistently capture objects of interest **remains undetermined**. This limitation **can be mitigated** using an image pyramid.

Computationally Heavy

- The sliding window technique necessitates a **full pass** on the full image. This can be **computationally expensive**. There is no mechanism to skip regions without helpful features.



(left) Aerial photo of houses.
To identify the swimming pool in the picture, the CNN has to slide through the whole image and its image pyramid.