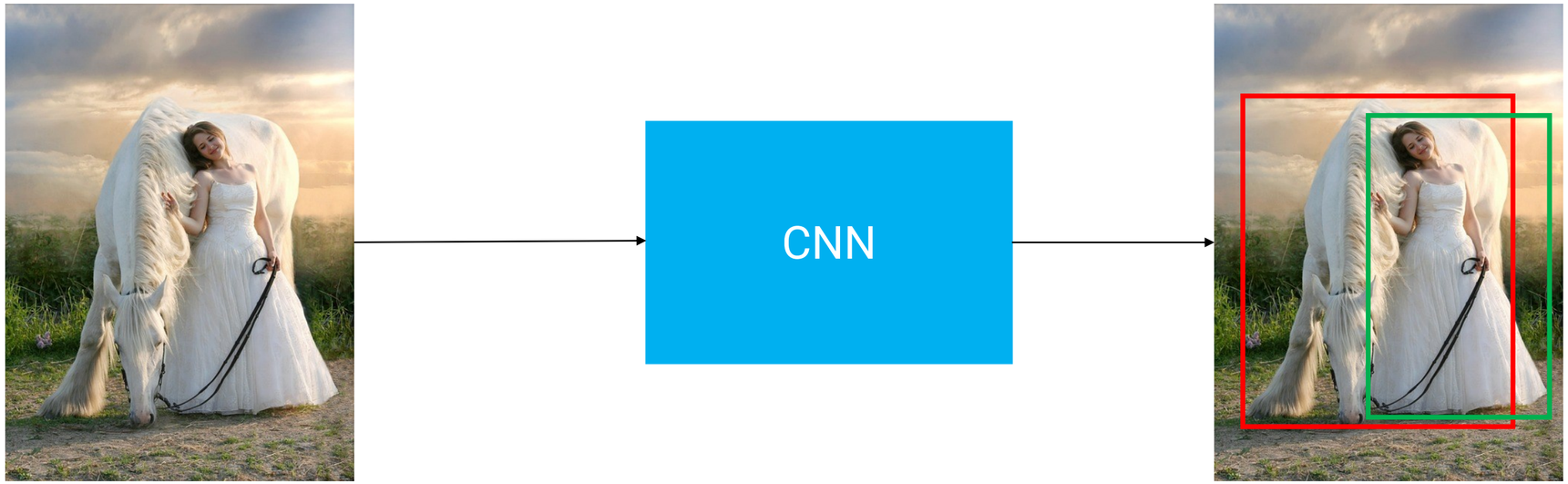
YOLO: You Only Look Once

Dr. Alireza Aghamohammadi

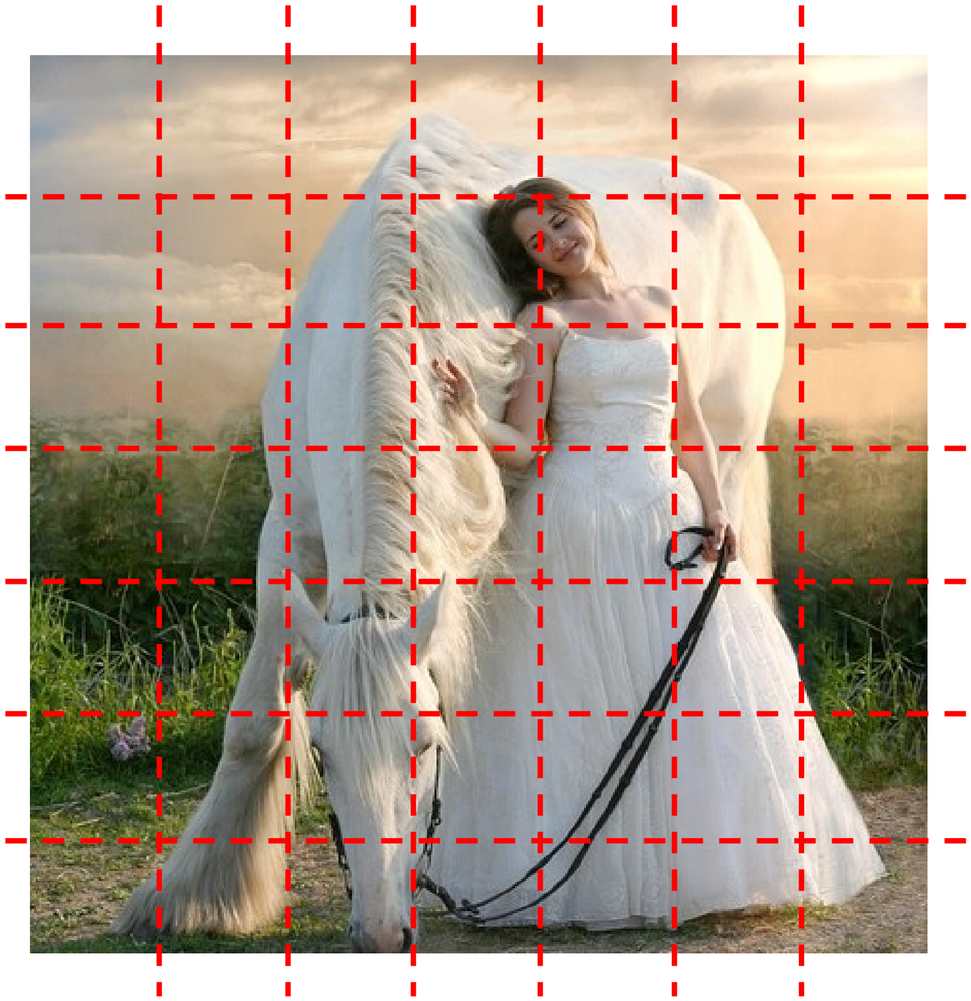
YOLO Idea

- Consider the task of object detection as a single regression problem that directly transforms image pixels into bounding box coordinates and class probabilities.



Steps in YOLO

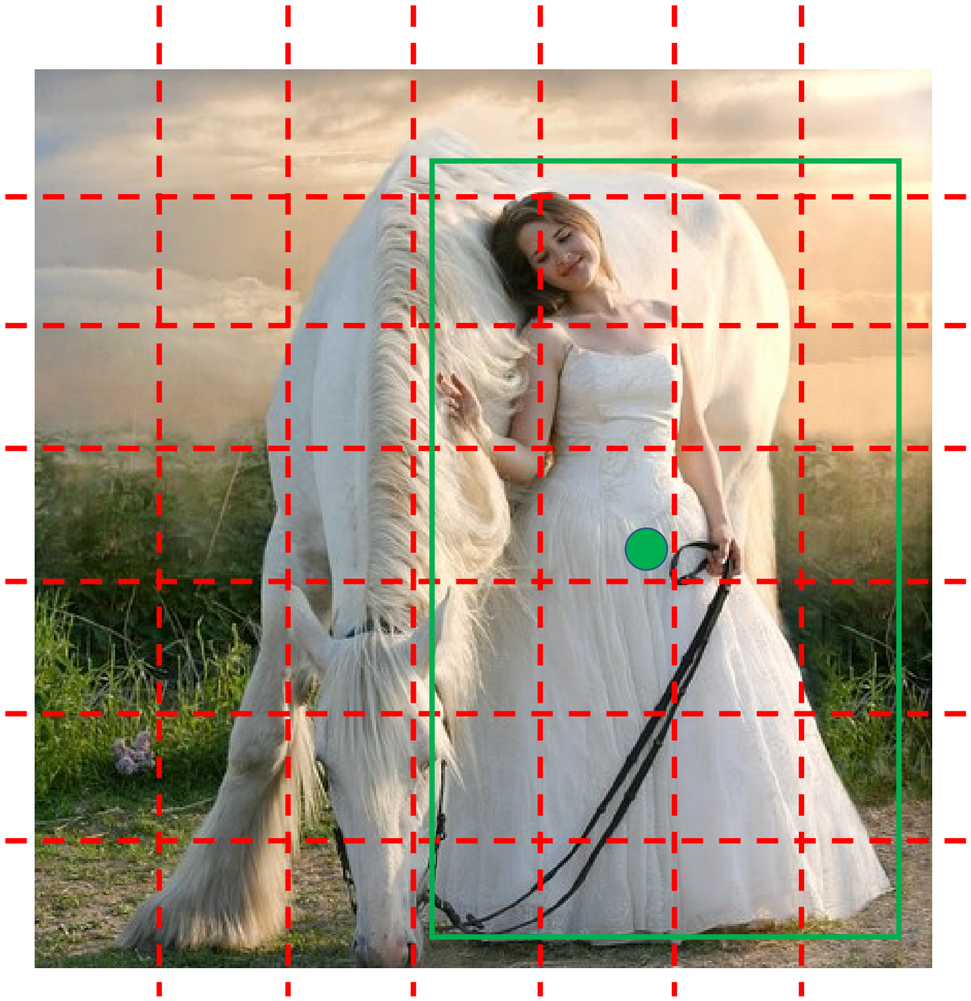
- Start with an input image (for instance, 1280x1748 pixels).
- Resize this image to a standard size of 448x448 pixels.
- Partition the image into a grid of SxS cells.
- Each individual cell in this grid will generate a prediction, which includes a bounding box.
- Output: $S \times S \times P$



Steps in YOLO

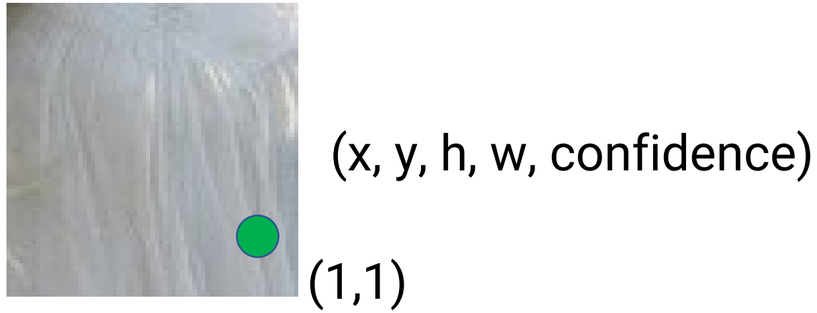
- If the center point of an object lands within a specific grid cell, then that particular cell takes on the responsibility of detecting that object.

(x, y, h, w, confidence)



Steps in YOLO

- Every cell prediction is composed of five numbers: x, y, w, h, and confidence.
- The (x, y) coordinates denote the center of the box in relation to the boundaries of the grid cell.
- The width (w) and height (h) are predicted in relation to the entire image.
- The confidence prediction signifies the Intersection over Union (IoU) between the predicted box and any actual ground truth box.



Architecture

Output: $S \times S \times (5 * B + C)$