VLMs on vision tasks

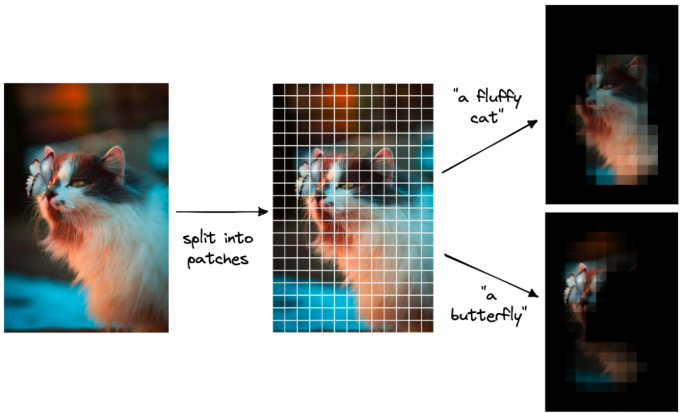

Figure 7: CLIP in object detection and localization