Hand-object interaction: a step towards action recognition

Abstract:

When dealing with a highly variable problem such as action recognition, focusing on a small area, such as
the hand's region, makes the problem more manageable, and enables us to invest relatively high amount
of resources needed for interpretation in a small but highly informative area of the image. In order to
detect this region of interest in the image and properly analyze it, I have built a process that includes
several steps, starting with a state of the art hand detector, incorporating both detection of the hand by
appearance and by estimation of human body pose. The hand detector is built upon a fully convolutional
neural network, detecting hands efficiently and accurately. The human body pose estimation starts with a
state of the art head detector and continues with a novel approach where each location in the image
votes for the position of each body keypoint, utilizing information from the whole image. Using dense,
multi-target votes enables us to compute image-dependent joint keypoint probabilities by looking at
consensus voting, and accurately estimates the body pose. Once the detection of the hands is complete,
an additional step of segmentation of the hand and fingers is made. In this step each hand pixel in the
image is labeled using a dense fully convolutional network. Finally, an additional step is made to segment
and identify the held object. Understanding the hand-object interaction is an important step toward
understanding the action taking place in the image. These steps enable us to perform fine interpretation
of hand-object interaction images as an essential step towards understanding the human-object
interaction and recognizing human activities.