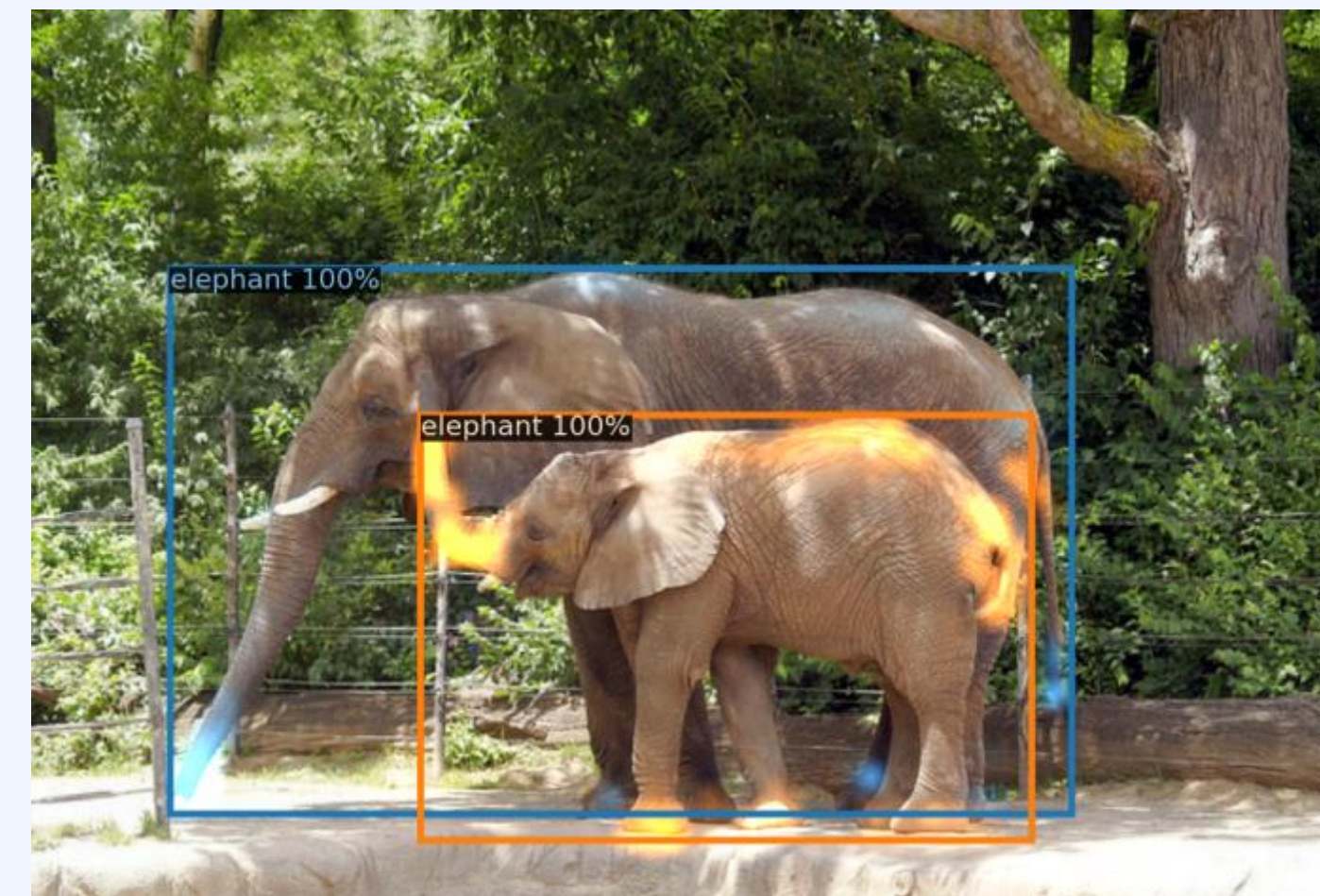
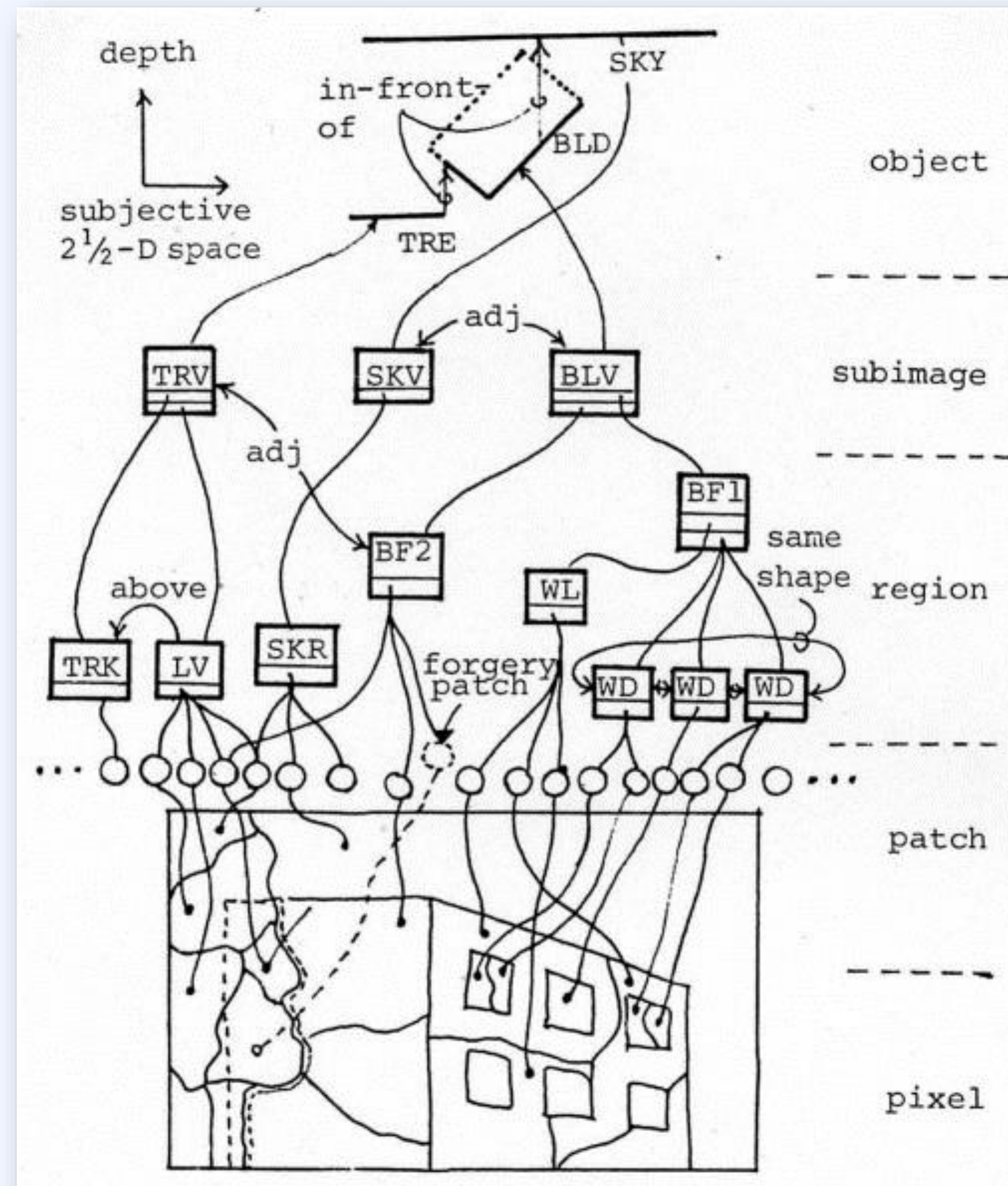


Scene Understanding

1 Overview

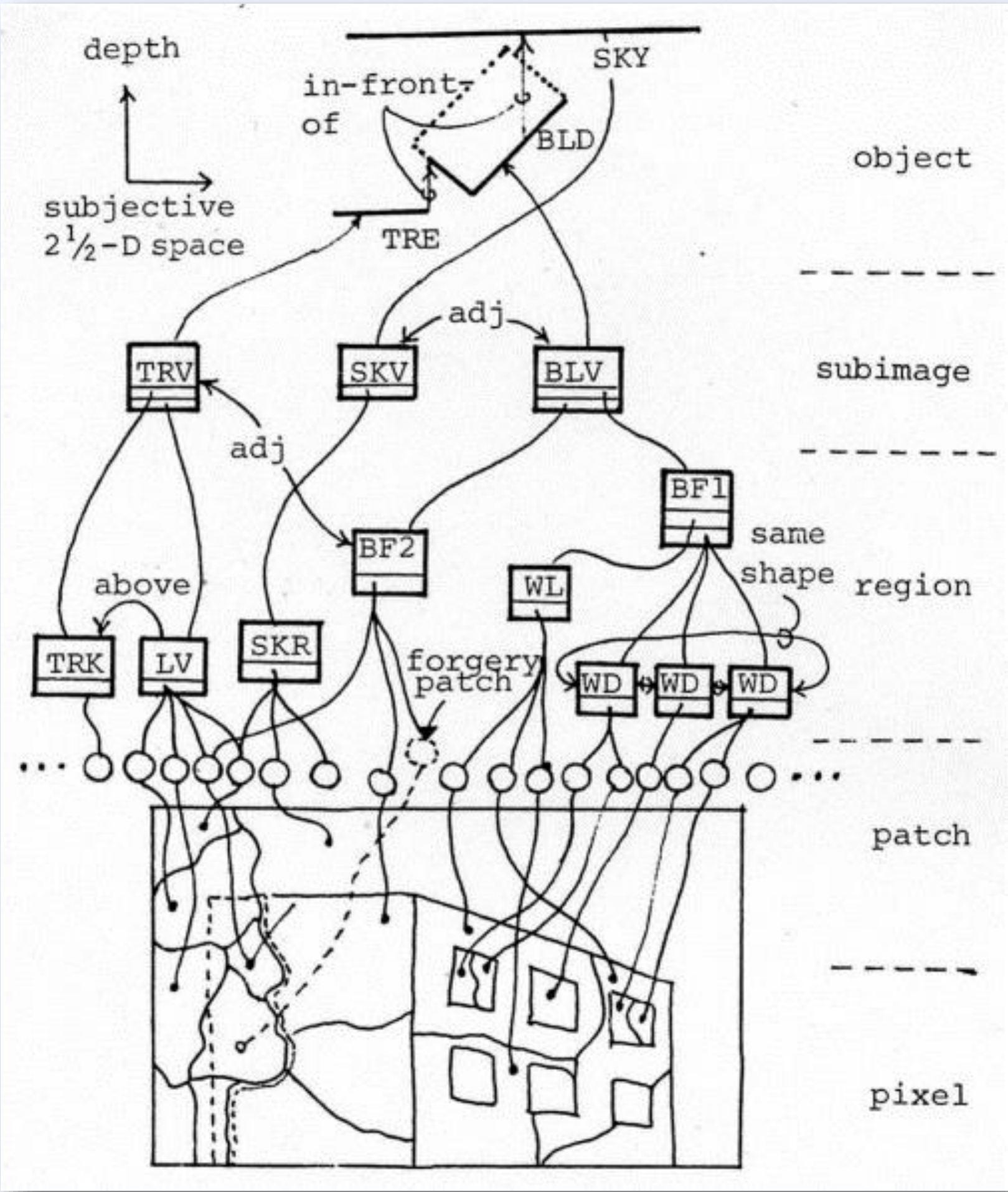
Scene Understanding Overview

1. Overview



Scene Understanding
Overview

1.
Overview



$(\text{bbox}_1^h, \text{bbox}_1^o, [\text{eat, hold}])$



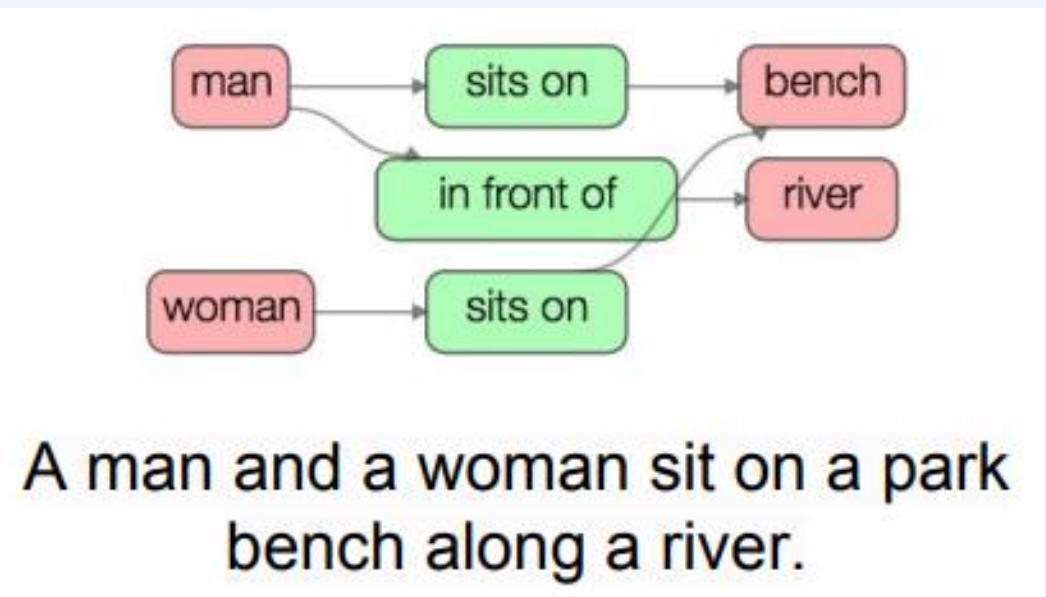
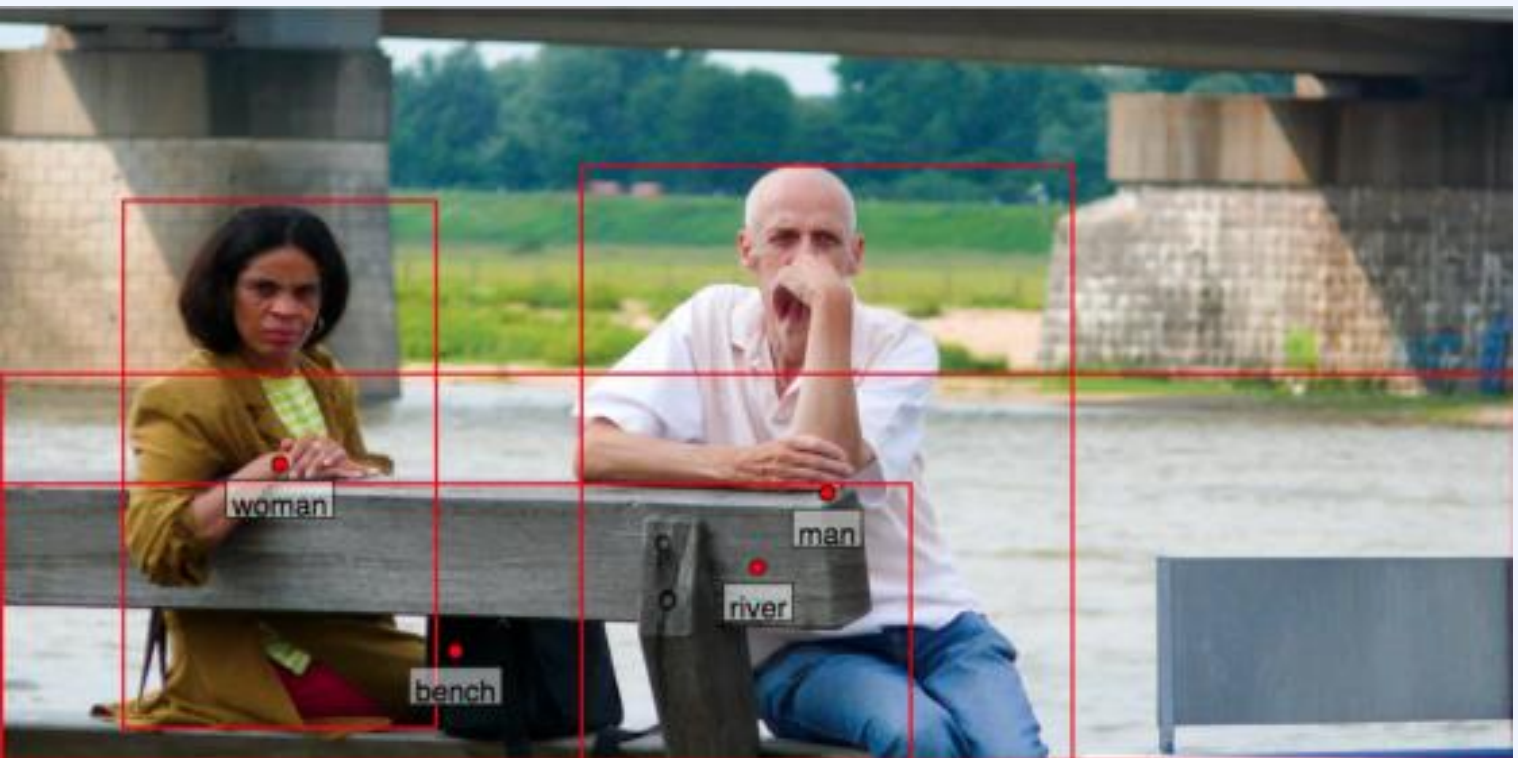
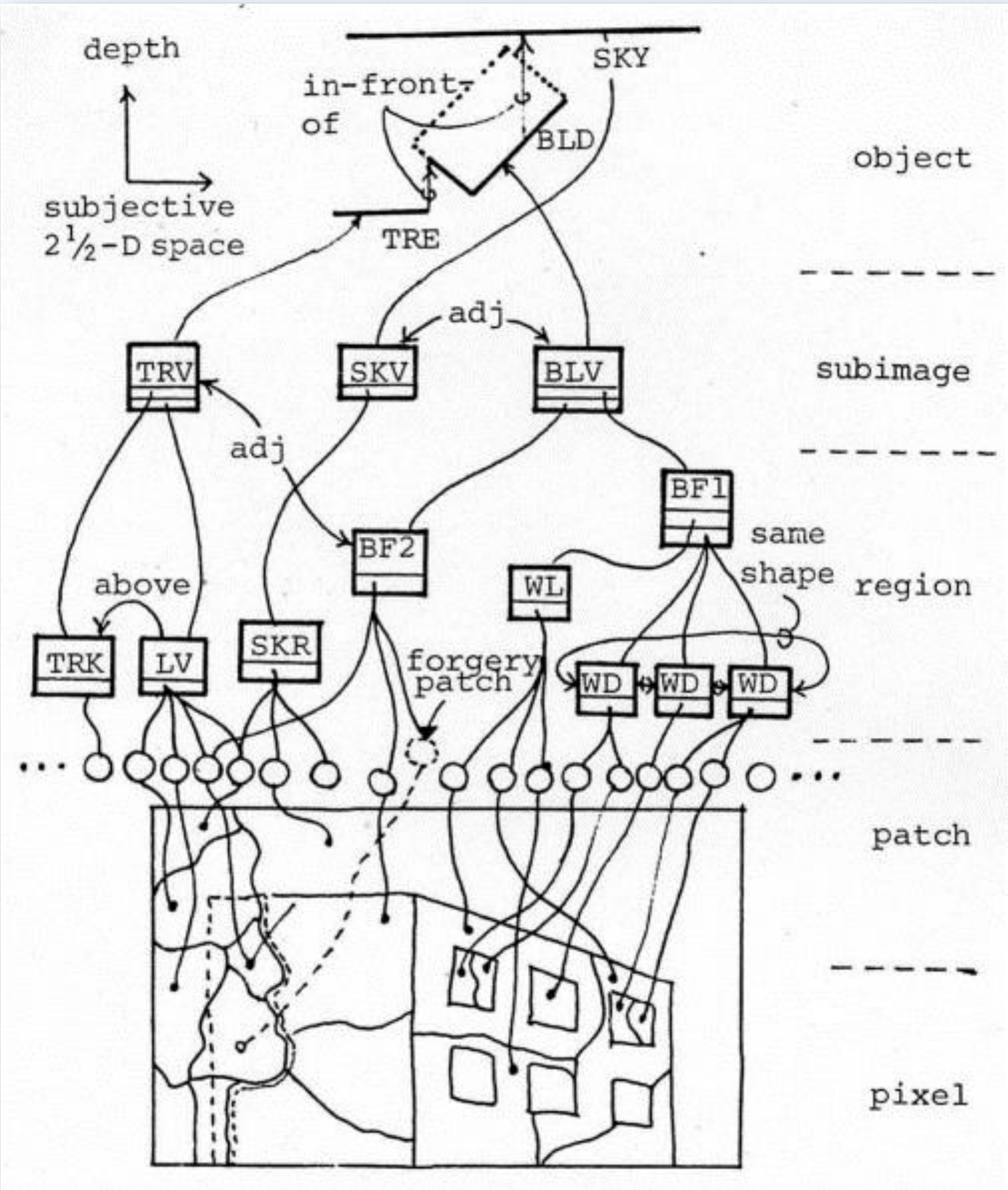
$(\text{bbox}_2^h, \text{bbox}_2^o, [\text{sit}])$



Scene Understanding Overview

1. Overview

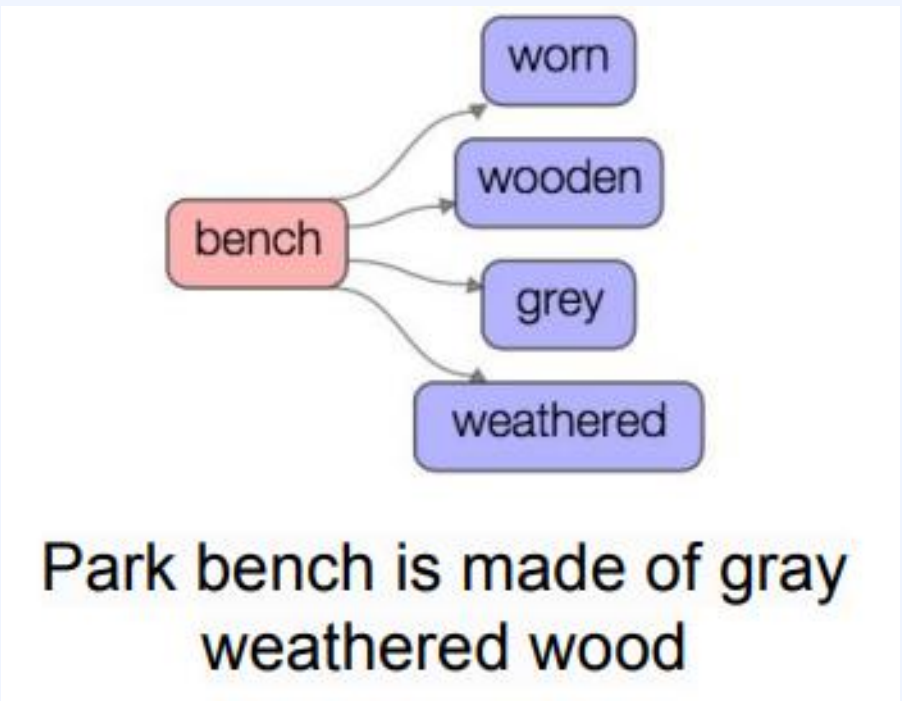
R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations



A man and a woman sit on a park bench along a river.



The man is almost bald



Park bench is made of gray weathered wood

Scene Understanding Overview

1.

Overview



Scene Understanding Overview

1.
Overview

Y. Zhang et al. Explicit Knowledge Incorporation for Visual Reasoning. CVPR

