

Input Image



Backbone



Feature extraction

Neck



Multi-resolution
feature agregation

Head



Generates final
predictions

