

Input Image

640 X 640

Backbone

Feature
Extraction

Neck

Multi-scale
Feature Fusion

Head

Classification &
Regression

Output Image

With bounding
boxes and classes