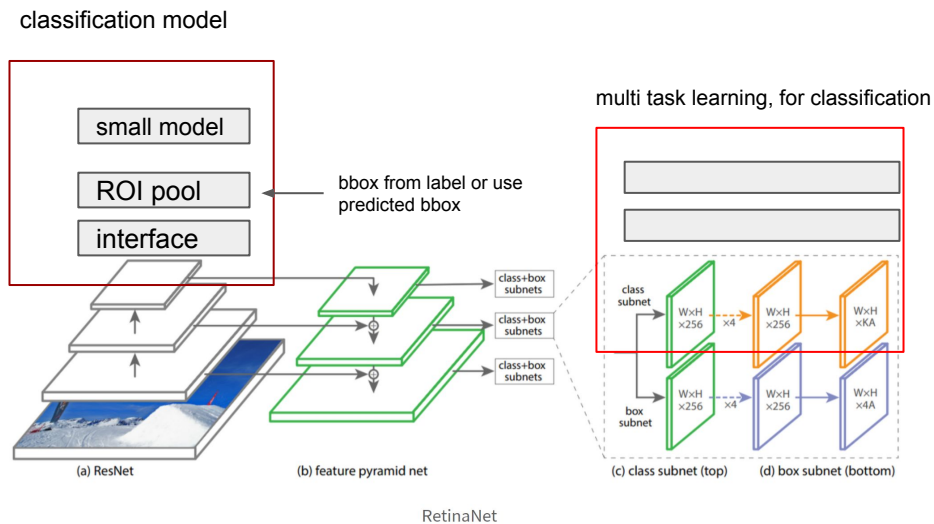


Proposed Architecture



method and key idea

- **detection model training**
freezing all detection model
training classification model
(use predicted bbox in inference)
- **advantage over legacy**
concentrate classification only(compare to detection only)
No dimension reduction(compare to detection only)
Use context (compare to detection/classification separate model)
- **For interface, dilate conv, sparse attn, swin attn**
try Roi pool or Roi align
shared block or separate block for multi task, and ratio test
need