

The model is based on the [2d implementation by matterport](#)

bb: bounding box

| Block name | Layer Type | Kernel Size | Activation | Filters | Block name | Layer Type | Kernel Size | Activation | Filters |
|----------------|------------|--|------------|---------------------|-------------|------------|---|------------|-------------------------|
| RPN class head | conv | 3 | relu | 512 | RPN bb head | conv | 3 | relu | 1024 |
| | Dropout | Rate: 0.2 | | | | conv | 1 | linear | # anchors per pixel * 6 |
| | conv | 1 | linear | # anchors per pixel | | reshape | output shape: (Batch, anchors per pixel * # pixels, 6) | | |
| | Reshape | output shape: (Batch, anchors per pixel * # pixels) | | | | | | | |
| | Activation | | sigmoid | | | | | | |

| | | | | | | | | | |
|-----------------|------------|-----------|---------|-----|--------------|---------|-------------------------------------|--------|---|
| RCNN Shared | conv | 7 | | 256 | | | | | |
| | batch norm | | | | | | | | |
| | activation | | relu | | | | | | |
| | conv | 1 | | 256 | | | | | |
| | batch norm | | | | | | | | |
| | activation | | relu | | | | | | |
| | dense | | | 512 | | | | | |
| RCNN class head | dropout | Rate: 0.5 | | | RCNN bb head | dense | | linear | 6 |
| | dense | | | 1 | | reshape | output shape: (Batch, # ROIs, 6) | | |
| | activation | | sigmoid | | | | | | |

| | | | | | |
|-----------|------------|-------------|------------|---------|--------|
| Mask head | Layer name | Kernel Size | Activation | Filters | stride |
| | conv | 3 | | 256 | 1 |
| | batch norm | | | | |
| | activation | | relu | | |
| | conv | 3 | | 256 | 1 |
| | batch norm | | | | |
| | activation | | relu | | |
| | conv | 3 | | 256 | 1 |
| | batch norm | | | | |
| | activation | | relu | | |
| | deconv | 3 | relu | 256 | 2 |
| | conv | 1 | sigmoid | 1 | |