

Instance Segmentation

Computer Vision



Why still use a two-stage object detector?

- Better recall of RPN as compared to SSD/YOLO
 - Trained with all object instances
 - o Generic first stage, usable for multi task
- Finer control over training classifier
 - Custom minibatch (sampling 3:1 negative samples)
- Instance-level multi task (Mask-RCNN)



Why still use a two-stage object detector?

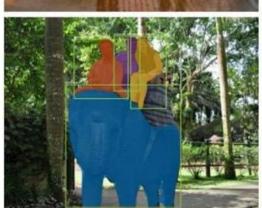
- Better recall of RPN as compared to SSD/YOLO
 - Trained with all object instances
 - o Generic first stage, usable for multi task
- Finer control over training classifier
 - Custom minibatch (sampling 3:1 negative samples)
- Instance-level multi task (Mask-RCNN)



Mask R-C NN – Towards Instance-Level

Understandi





This file is meant for personal use by rg.ravigupta91@gmail.com only.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibering or publishing the contents in part or full is liable for legal action.

greatlearningMask R-CNN – Towards Instance-Level Understanding



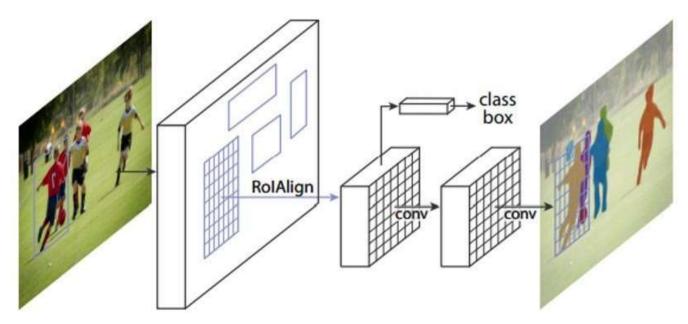
Zoom in on instances

This file is meant for personal use by rg.ravigupta91@gmail.com only. Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution problem or publishing the contents in part or full is liable for legal action.



Mask R-CNN

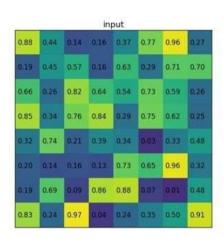
Preserves pixel-to-pixel alignment

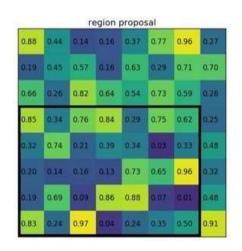


This file is meant for personal use by rg.ravigupta91@gmail.com only. Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution problem or publishing the contents in part or full is liable for legal action.

greatlearning

Quantization – loss of pixel-to-pixel alignment





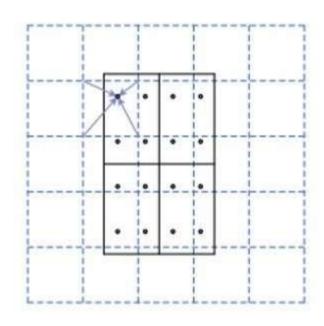
| pooling sections | | | | | | | |
|------------------|------|------|------|------|------|------|------|
| 0.88 | 0.44 | 0.14 | 0.26 | 0.37 | 0.77 | 0.96 | 0.27 |
| 0.19 | 0.45 | 0.57 | 0.16 | 0.63 | 0.29 | 0.71 | 0.70 |
| 0.66 | 0.26 | 0.82 | 0.64 | 0.54 | 0.73 | 0.59 | 0.26 |
| 0.85 | 0.34 | 0.76 | 0.84 | 0.29 | 0.75 | 0.62 | 0.25 |
| 0.32 | 0.74 | 0.21 | 0.39 | 0.34 | 0.03 | 0,33 | 0.48 |
| 0.20 | 0.14 | 016 | 0.13 | 0.73 | 0.65 | 0.96 | 0.32 |
| 0.19 | 0.69 | 0.09 | 0.86 | 0.88 | 0.07 | 0.01 | 0.48 |
| 0.83 | 0.24 | 0.97 | 0.04 | 0.24 | 0.35 | 0.50 | 0.91 |





Rol Align - Improvement on Rol Pooling

- Input: Feature map (5x5 here) and region proposal (normalized float coordinates)
- Output: 2x2 'pooled' bins
- Sample 4 points in every bin uniformly
- Compute value at each bin using bilinear interpolation
- Max or average the 4 bins





Class Imbalance in Training a Classifier

- While training detectors, maximum samples are background (negatives)
- Faster R -C NN: Ratio of 3 negatives to 1 positive is maintained while training classifier head Custom minibatch
- Not easy in single stage detectors



Class Imbalance in Training a Classifier

Cross entropy loss

$$CE(p_t) = -\log(p_t)$$

Balanced cross entropy loss

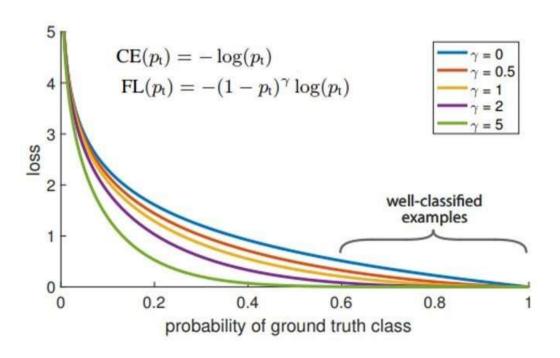
$$CE(p_t) = -\alpha_t \log(p_t)$$

• Focal

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t)$$



Focal Loss

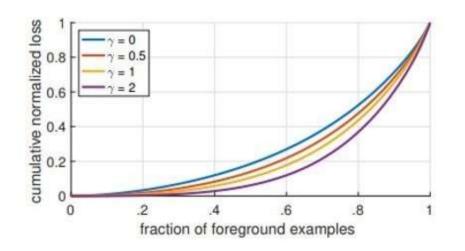


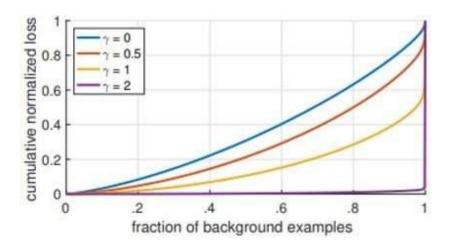
This file is meant for personal use by rg.ravigupta91@gmail.com only. Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution probleming or publishing the contents in part or full is liable for legal action.



Focal Loss

$$FL(p_{t}) = -(1 - p_{t})^{\gamma} \log(p_{t})$$







Thank you!

Happy Learning:)