2020.09.09

Focal Loss

Tsung-Yi Lin Priya Goyal Ross Girshick Kaiming He Piotr
Dollar´ Facebook AI Research (FAIR)

20151739 공대현

❖ Focal Loss의 필요성

- RCNN 계열의 Two-stage Network와 SSD,YOLO 같은 One-stage Network의 Accuracy 차이가 존재함

- Object Detection에서 Class의 Imbalance가 있기 때문 (Background↑,Foreground↓)

- Two-stage Network 는 Selective Search를 통해 Class Imbalance를 해결 (ROI:2000개), Background : Foreground = 3:1

- 하지만 One-stage Network는 ROI로 Background Examples가 매우 많이 나옴(“Overwhelm”)

→ Focal Loss 를 이용해 Easy example(Backgrounds)에 대한 Loss 비중을 낮추고, Hard example(Fore Ground)에 대한 Loss 비중을 높임

❖ Focal Loss 식

- Cross Entropy Loss :

$$\mathrm{CE}(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise.} \end{cases}$$

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise,} \end{cases}$$

$$\mathrm{CE}(p, y) = \mathrm{CE}(p_t) = -\log(p_t).$$

- Focal Loss :

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

❖ Focal Loss 식

$P_1 = 0.9$, $P_2 = 0.1$ $(1-P_1)$

CE Loss 합 : $-\sum_{c=1}^{2} L_c \log P_c$

i) if $L = [1, 0]$

CE : $-1 \cdot \underbrace{\log 0.9}_{-0.09} + 0 \cdot \underbrace{\log 0.1}_{-1}$

$= -1 \cdot -0.09 + 0 \cdot (-1) = \boxed{0.09}$

good predict

ii) if $L = [0, 1]$

CE : $0 \cdot \underbrace{\log 0.9}_{-0.09} + -1 \cdot \underbrace{\log 0.1}_{-1}$

$= 0 \cdot -0.09 + -1 \cdot (-1) = \boxed{1}$

bad predict

$P_1 = 0.9$, $P_2 = 0.1$ $(1-P_1)$

$-\log(P_t)$

$\downarrow$

Focal Loss $(\alpha = 1, \gamma = 2)$: $-(1-P_t)^{\gamma} \log(P_t)$

i) if $L = [1, 0]$

FL : $-1 \cdot (1-0.9)^2 \cdot \underbrace{\log 0.9}_{-0.09} + 0 \cdot \underbrace{\log 0.1}_{-1}$

$= -1 \cdot 0.1^2 (-0.09) + 0 \cdot (-1) = \boxed{0.0729}$

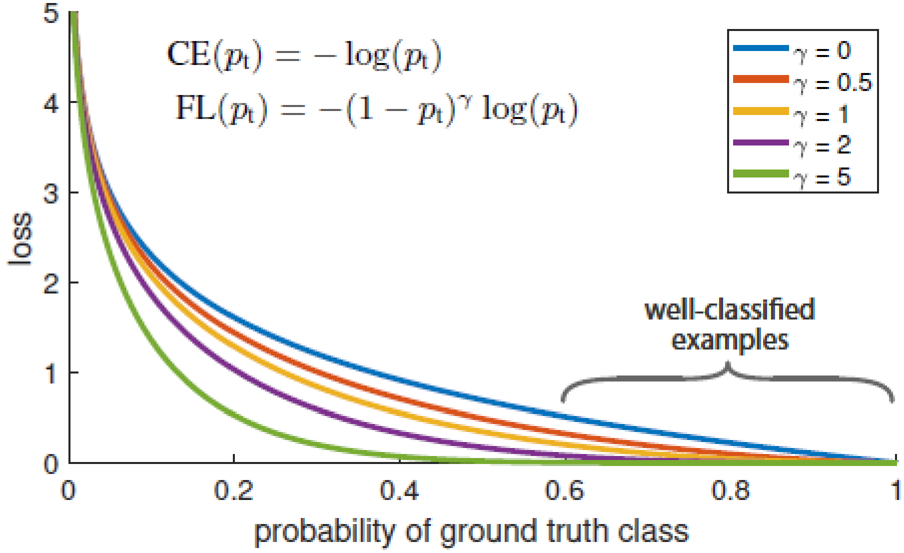
good predict

ii) if $L = [0, 1]$

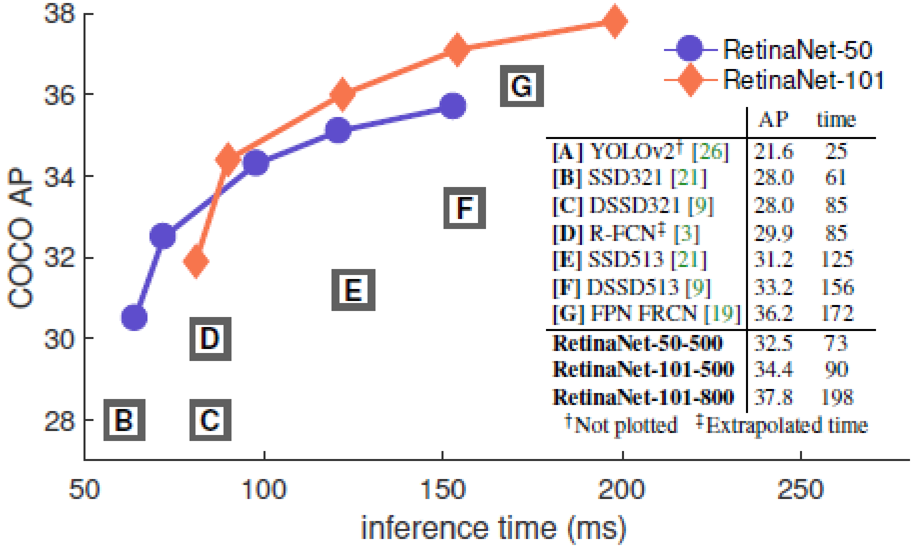
CE : $0 \cdot \underbrace{\log 0.9}_{-0.09} + -1 \cdot 0.9^2 \underbrace{\log 0.1}_{-1}$

$= 0 \cdot -0.09 + 0.9^2 \cdot (-1) \cdot (-1) = \boxed{0.81}$

bad predict



$\mathrm{CE}(p_t) = -\log(p_t)$
$\mathrm{FL}(p_t) = -(1-p_t)^\gamma \log(p_t)$
γ = 0
γ = 0.5
γ = 1
γ = 2
γ = 5
well-classified examples
loss
probability of ground truth class
0
0.2
0.4
0.6
0.8
1
1
2
3
4
5



RetinaNet-50
RetinaNet-101
COCO AP
inference time (ms)
28
30
32
34
36
38
50
100
150
200
250
B
C
D
E
F
G
AP time
[A] YOLOv2† [26] 21.6 25
[B] SSD321 [21] 28.0 61
[C] DSSD321 [9] 28.0 85
[D] R-FCN‡ [3] 29.9 85
[E] SSD513 [21] 31.2 125
[F] DSSD513 [9] 33.2 156
[G] FPN FRCN [19] 36.2 172
RetinaNet-50-500 32.5 73
RetinaNet-101-500 34.4 90
RetinaNet-101-800 37.8 198
†Not plotted ‡Extrapolated time

α	AP	AP_{50}	AP_{75}
.10	0.0	0.0	0.0
.25	10.8	16.0	11.7
.50	30.2	46.7	32.8
.75	31.1	49.4	33.0
.90	30.8	49.7	32.3
.99	28.7	47.4	29.9
.999	25.1	41.7	26.1

(a) **Varying α for CE loss** ($\gamma = 0$)

γ	α	AP	AP_{50}	AP_{75}
0	.75	31.1	49.4	33.0
0.1	.75	31.4	49.9	33.1
0.2	.75	31.9	50.7	33.4
0.5	.50	32.9	51.7	35.2
1.0	.25	33.7	52.0	36.2
2.0	.25	**34.0**	**52.5**	**36.5**
5.0	.25	32.2	49.6	34.8

(b) **Varying γ for FL** (w. optimal α)

#sc	#ar	AP	AP_{50}	AP_{75}
1	1	30.3	49.0	31.8
2	1	31.9	50.0	34.0
3	1	31.8	49.4	33.7
1	3	32.4	52.3	33.9
2	3	**34.2**	**53.1**	**36.5**
3	3	34.0	52.5	**36.5**
4	3	33.8	52.1	36.2

(c) **Varying anchor scales and aspects**

- 매우 불균형한 데이터셋 Training : ex) Object Detection, Face Recognition ..