

라벨링 문제

1. Annotation Error의 영향

(<https://understand.ai/blog/annotation/machine-learning/autonomous-driving/2021/06/01/impact-of-annotation-errors-on-neural-networks.html>)

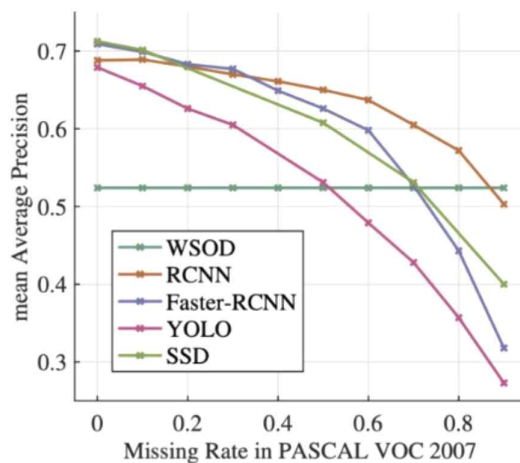
[1] **Incorrect class**: An object is classified incorrectly, e.g. a vehicle is labeled as pedestrian.

[2] **Incorrect attribute**: The state of an object is not described correctly, e.g. a car in motion is labeled as parked.



Ground truth on the left, small noise (0.08) in the middle and large noise (0.13) on the right. Source: VOC dataset

[3] **Missing annotation**: An object is not annotated even though it should be.
(false negative를 학습하므로 학습에 악영향을 준다)

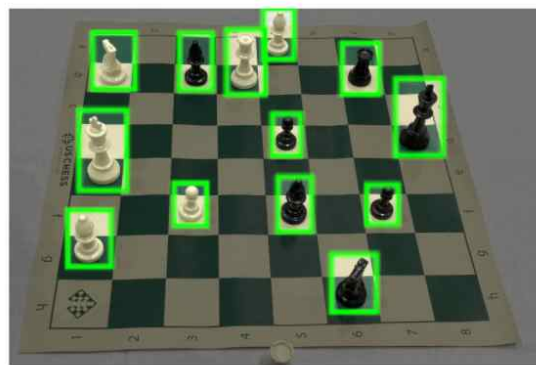
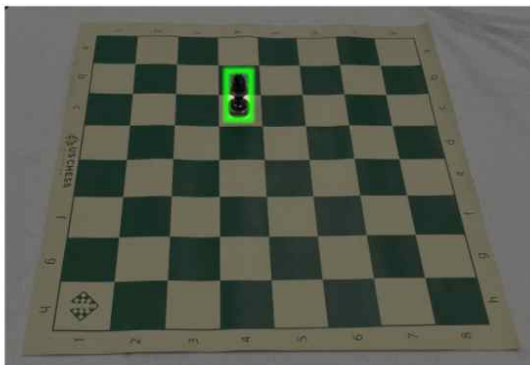


The Mean Average Precision (mAP) for one WSOD detector and four classical FSOD detectors training under different instance-level missing label rates (Mr=0.0:1.0:0.9) of the training data set [Xu et al., 2019]

[4] **Redundant annotation**: An object is annotated even though it shouldn't be

[5] **Incorrect annotation size or position**: An object is not annotated precisely enough, not fitting to its actual dimensions.

[6] **Incorrect annotation position:** An object is not annotated precisely enough, not placed at its actual position.



Four sample chess images. [Dataset source](#).

(<https://blog.roboflow.com/missing-and-null-image-annotations/>)

2. Annotation 생략 문제:

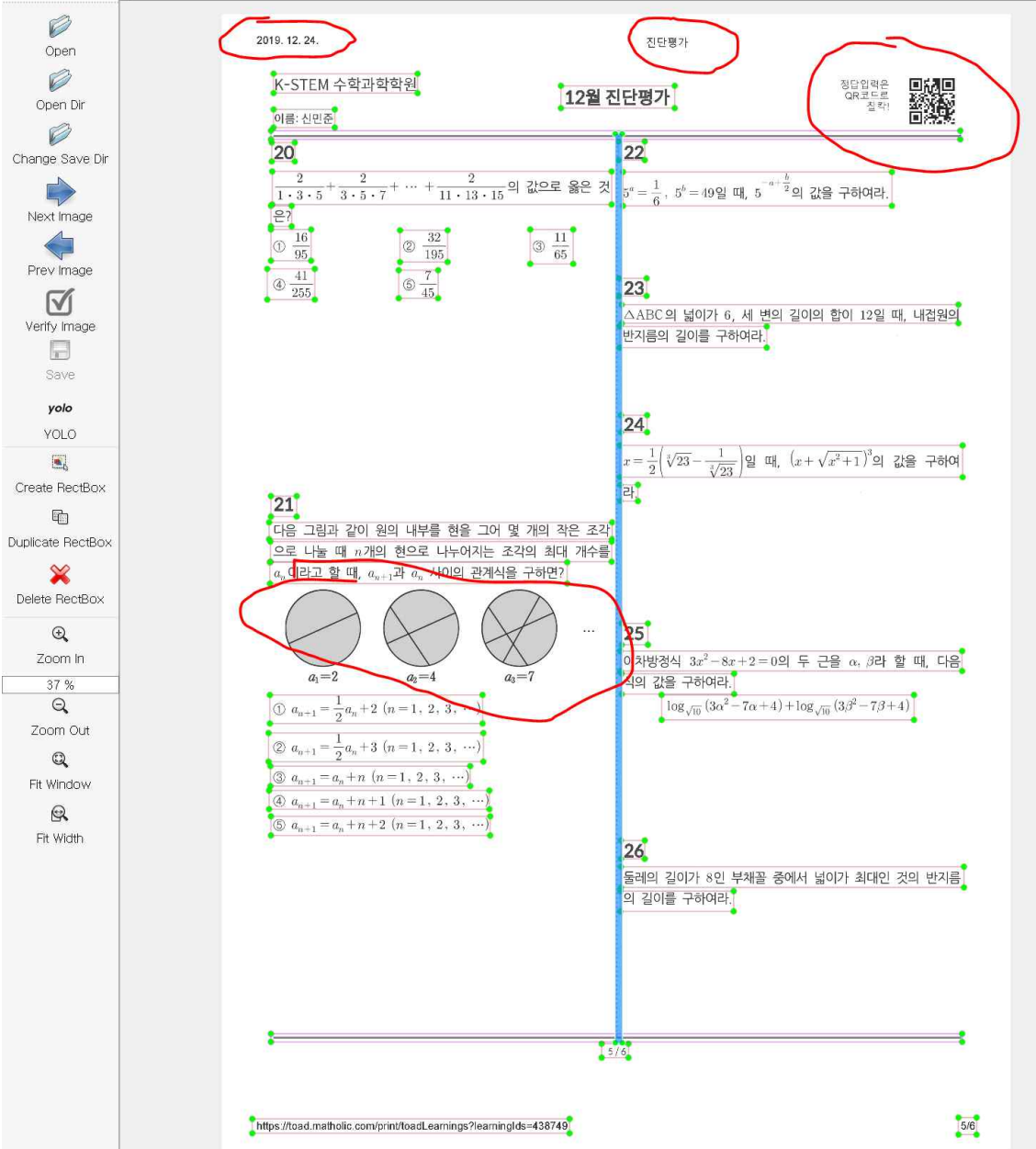
학습시에, 실재하는 특정 Object를 Object가 아닌 배경(background)으로 학습하라는 잘못된(혼동된) 명령을 신경망에게 주어서 학습이 불안정하게(그런 경우가 많을 시에는 학습이 안되게) 한다.

특히, 학습 이미지에 반복적으로 나타나는 Object일수록 더욱 유의하여야 한다 (샘플 수가 많을 테니)

MisLabeling 예 - Annotation 생략

labellmg D:\myprojects\datasets\document\images\train\Wh_1_p_5.jpg [7 / 377]

File Edit View Help



2019. 12. 24.

K-STEM 수학과학학원

이름: 신민준

진단평가

정답입력은
QR코드부터
철각!



12월 진단평가

20

$\frac{2}{1 \cdot 3 \cdot 5} + \frac{2}{3 \cdot 5 \cdot 7} + \dots + \frac{2}{11 \cdot 13 \cdot 15}$ 의 값으로 옳은 것은?

① $\frac{16}{255}$

② $\frac{32}{195}$

③ $\frac{11}{65}$

④ $\frac{41}{255}$

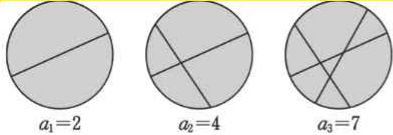
⑤ $\frac{7}{45}$

22

$5^a = \frac{1}{6}$, $5^b = 49$ 일 때, $5^{-a+\frac{b}{2}}$ 의 값을 구하여라.

21

다음 그림과 같이 원의 내부를 현을 그어 몇 개의 작은 조각으로 나눌 때 n 개의 현으로 나누어지는 조각의 최대 개수를 a_n 이라고 할 때, a_{n+1} 과 a_n 사이의 관계식을 구하면?



① $a_{n+1} = \frac{1}{2}a_n + 2$ ($n=1, 2, 3, \dots$)

② $a_{n+1} = \frac{1}{2}a_n + 3$ ($n=1, 2, 3, \dots$)

③ $a_{n+1} = a_n + n$ ($n=1, 2, 3, \dots$)

④ $a_{n+1} = a_n + n + 1$ ($n=1, 2, 3, \dots$)

⑤ $a_{n+1} = a_n + n + 2$ ($n=1, 2, 3, \dots$)

23

△ABC의 넓이가 6, 세 변의 길이의 합이 12일 때, 내접원의 반지름의 길이를 구하여라

24

$x = \frac{1}{2} \left(\sqrt[3]{23} - \frac{1}{\sqrt[3]{23}} \right)$ 일 때, $(x + \sqrt{x^2 + 1})^3$ 의 값을 구하여라.

25

이차방정식 $3x^2 - 8x + 2 = 0$ 의 두 근을 α, β 라 할 때, 다음 식의 값을 구하여라

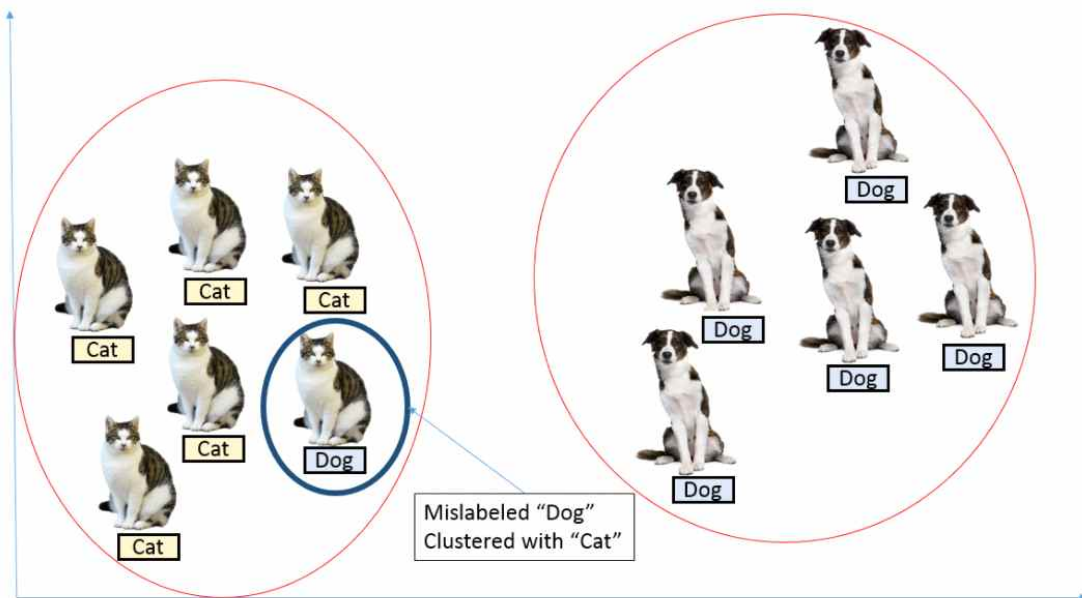
$\log_{\sqrt{10}}(3\alpha^2 - 7\alpha + 4) + \log_{\sqrt{10}}(3\beta^2 - 7\beta + 4)$

26

둘레의 길이가 8인 부채꼴 중에서 넓이가 최대인 것의 반지름의 길이를 구하여라.

3. Label Smoothing : 라벨링 오류에 대한 대책

학습시에,



```
tf.losses.softmax_cross_entropy(  
    onehot_labels,  
    logits,  
    weights=1.0,  
    label_smoothing=0,  
    scope=None,  
    loss_collection=tf.GraphKeys.LOSSES,  
    reduction=Reduction.SUM_BY_NONZERO_WEIGHTS  
)
```

arguments in softmax cross entropy loss

If `label_smoothing` is nonzero, smooth the labels towards $1/\text{num_classes}$:

$$\text{new_onehot_labels} = \text{onehot_labels} * (1 - \text{label_smoothing}) + \text{label_smoothing} / \text{num_classes}$$

What does this mean?

Well, say you were training a model for binary classification. Your labels would be 0 — cat, 1 — not cat.

Now, say you `label_smoothing = 0.2`

Using the equation above, we get:

$$\text{new_onehot_labels} = [0 \ 1] * (1 - 0.2) + 0.2 / 2 = [0 \ 1] * (0.8) + 0.1$$

$$\text{new_onehot_labels} = [0.9 \ 0.1]$$