Faster R-CNN Real-Time Object Detection with Region Proposal Networks

1st Paper Study | 2022.07.06 Wed.

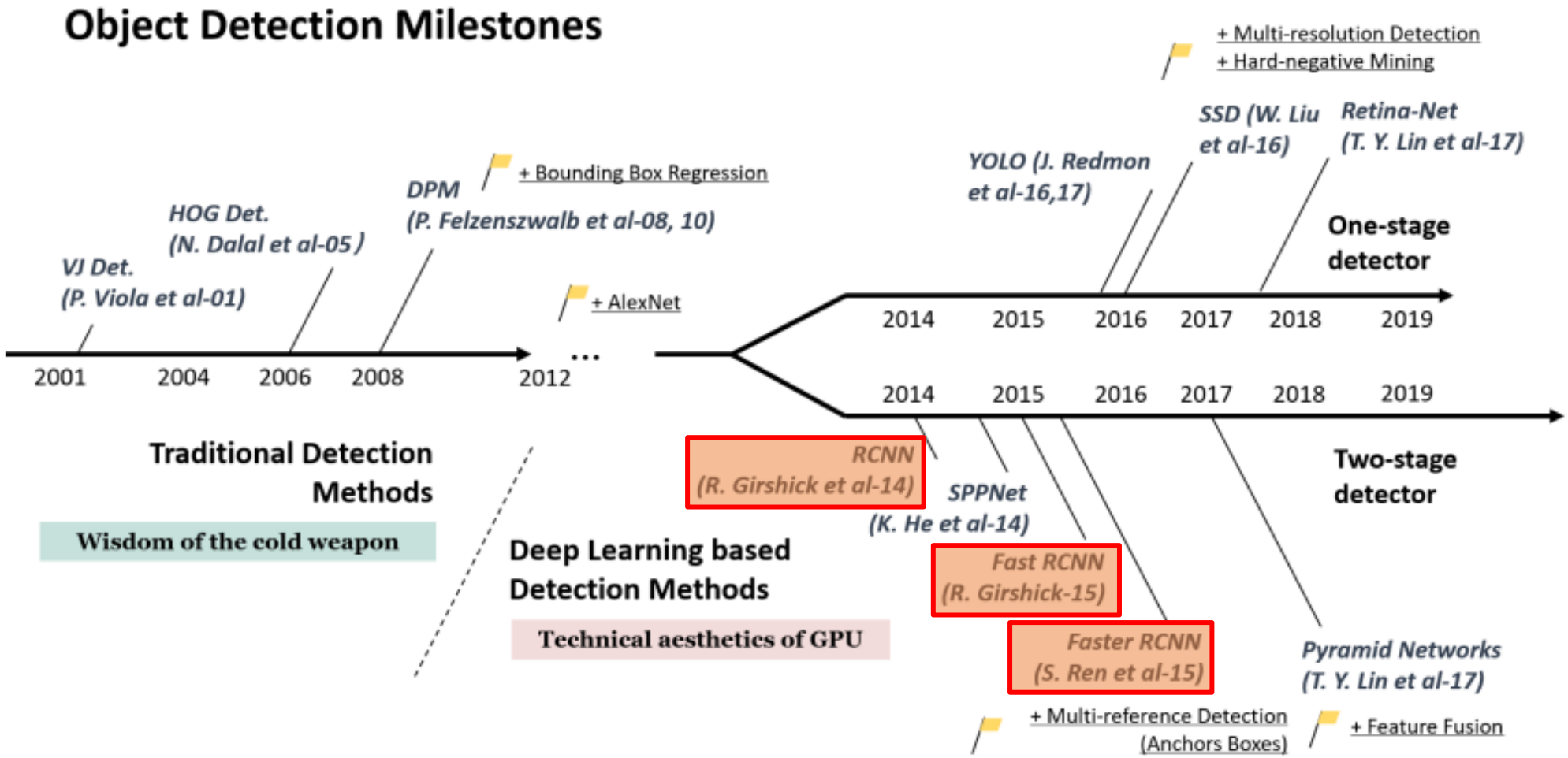
홍 세 현 Hong Sehyun

STRADVISION

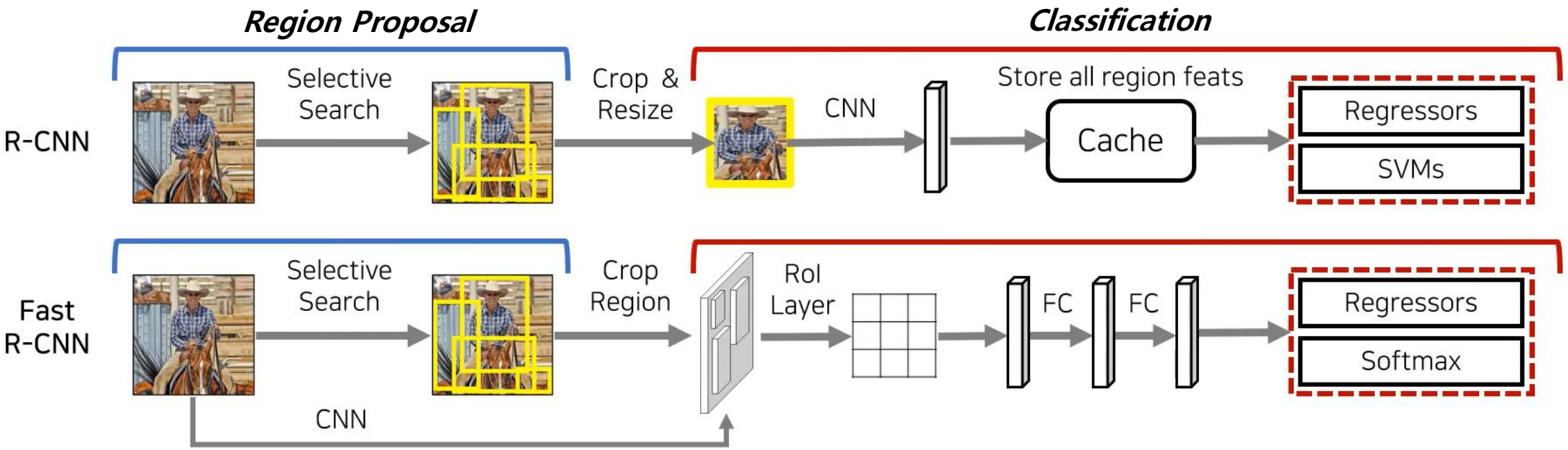
Outline

1. Introduction
2. Faster R-CNN
3. Experiments & Result
4. Reference
5. Q & A

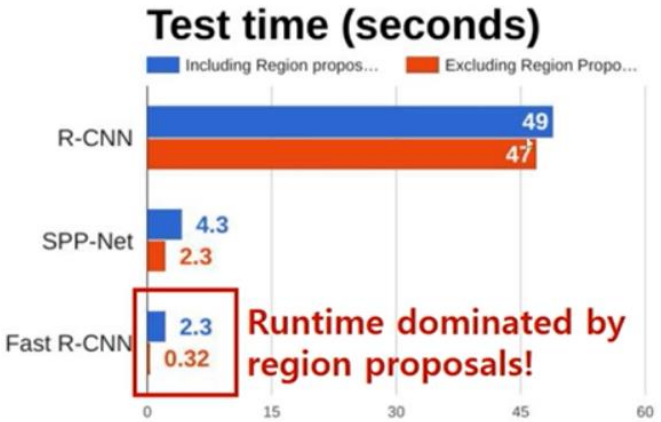
Introduction



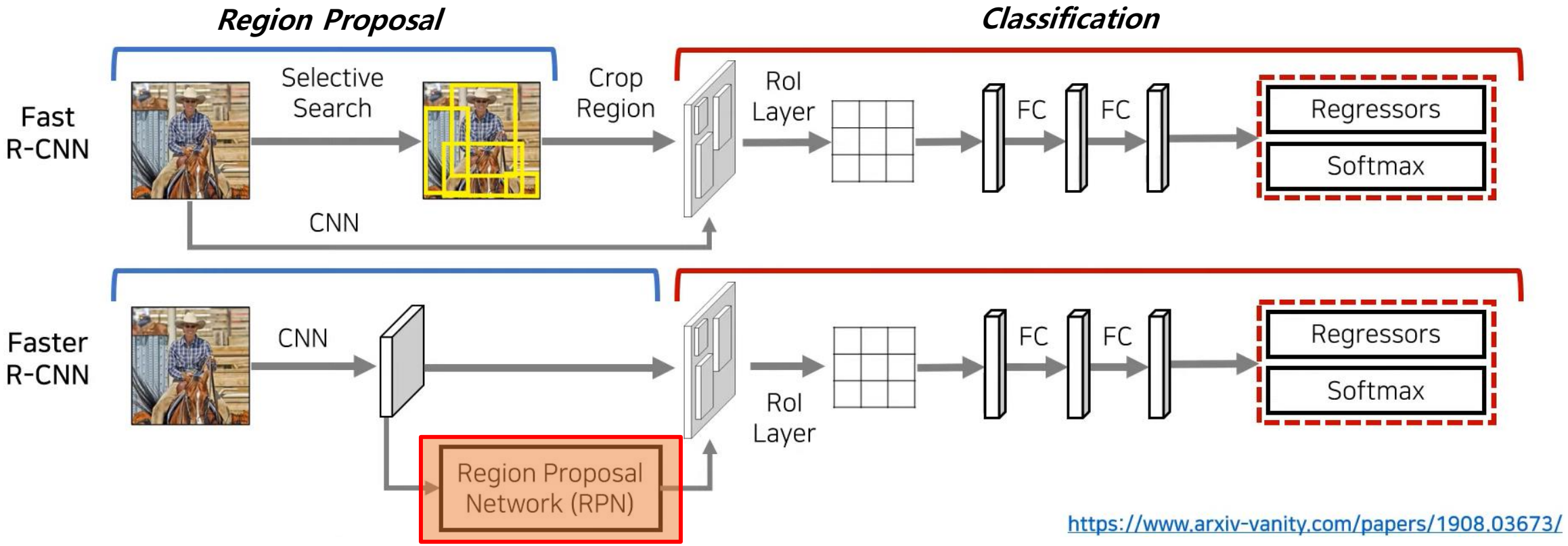
Introduction



Disadvantages of Existing Models

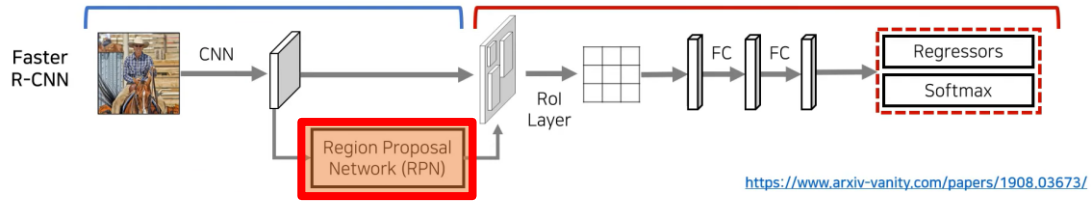


Faster R-CNN



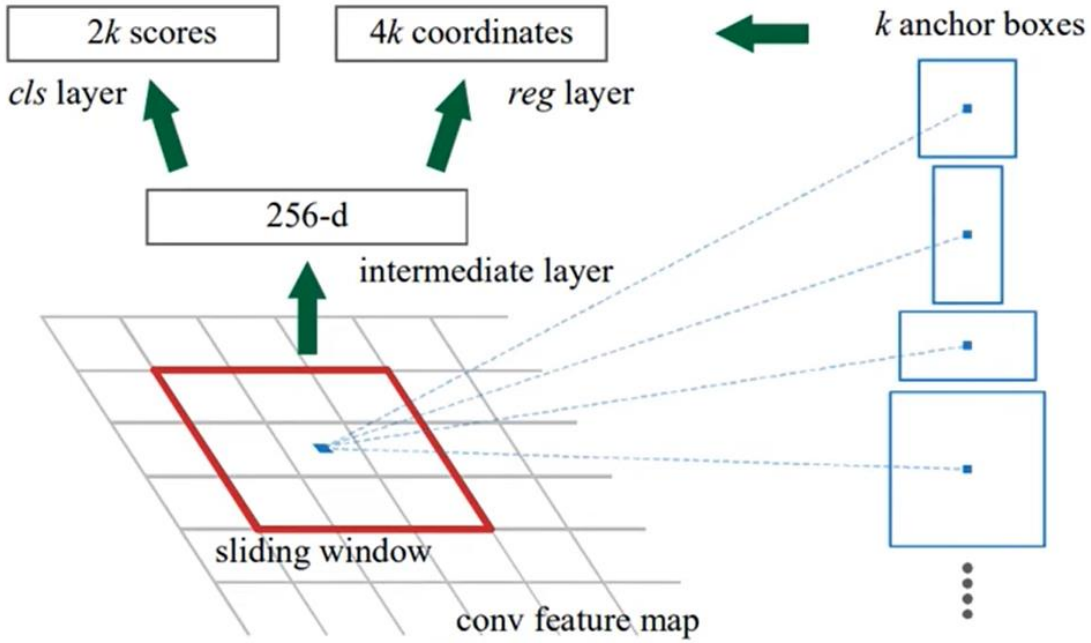
https://www.arxiv-vanity.com/papers/1908.03673/

Faster R-CNN

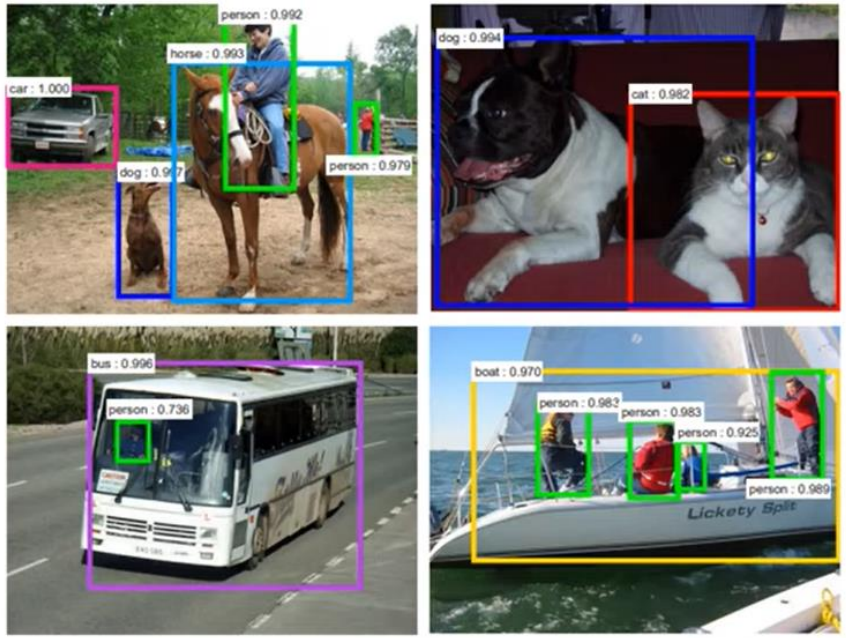


Classify
Obj. / NOT-obj.

Regression
Tuning Box Locations



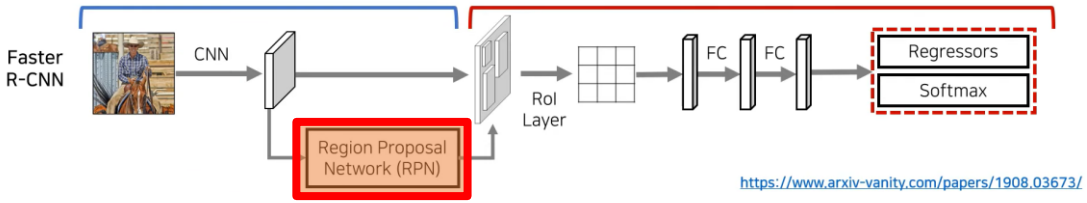
Regression gives offsets from anchor boxes



Using *k anchor Boxes* at each Location

- **3 scale (128x128, 256x256, 512x512)**
- **3 ratio (2:1, 1:1, 1:2)**

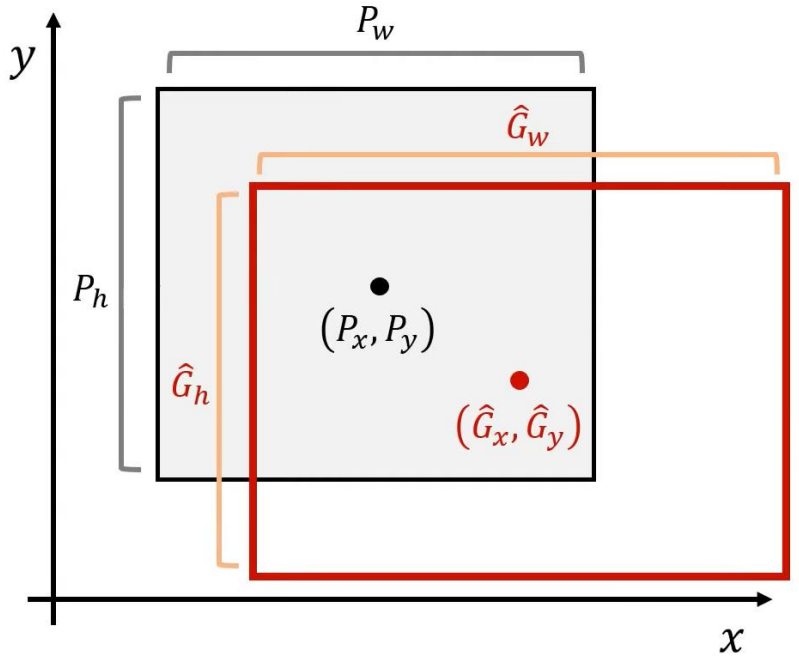
Faster R-CNN



Learning Dataset $\{(P^i, G^i)\}_{i=1,\ldots,N}$

Predicted Location $P^i = (P_x^i, P_y^i, P_w^i, P_h^i)$

Label Location $G^i = (G_x^i, G_y^i, G_w^i, G_h^i)$



Bounding Box Linear Regression

$$\hat{G}_x = P_w d_x(P) + P_x$$

$$\hat{G}_y = P_h d_y(P) + P_y$$

$$\hat{G}_w = P_w \exp(d_w(P))$$

$$\hat{G}_h = P_h \exp(d_h(P))$$

Faster R-CNN

Intersection Over Union(*IoU*) = 두 Bounding Box가 겹치는 비율

- **성능 평가** : mAP@0.5는 정답과 예측의 IoU가 50% 이상일 때 정답으로 판정하겠다는 의미

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

(교집합)

(합집합)

```
If, IoU > 0.7           : 1(Positive)
If, IoU < 0.3           : -1(Negative)
If, 0.3 <= IoU <= 0.7   : 0(Background, Ignore)
```

Faster R-CNN

Intersection Over Union(*IoU*) = **두 Bounding Box가 겹치는 비율**

Loss Function

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*).$$

$$t_x = (x - x_a)/w_a, \quad t_y = (y - y_a)/h_a,$$
$$t_w = \log(w/w_a), \quad t_h = \log(h/h_a),$$
$$t_x^* = (x^* - x_a)/w_a, \quad t_y^* = (y^* - y_a)/h_a,$$
$$t_w^* = \log(w^*/w_a), \quad t_h^* = \log(h^*/h_a),$$

p_i^* = Class Label

t_i^* = $t_x\ t_y\ t_w\ t_h$
= Ground Truth Bounding Box의 정보를 가지는 tuple data

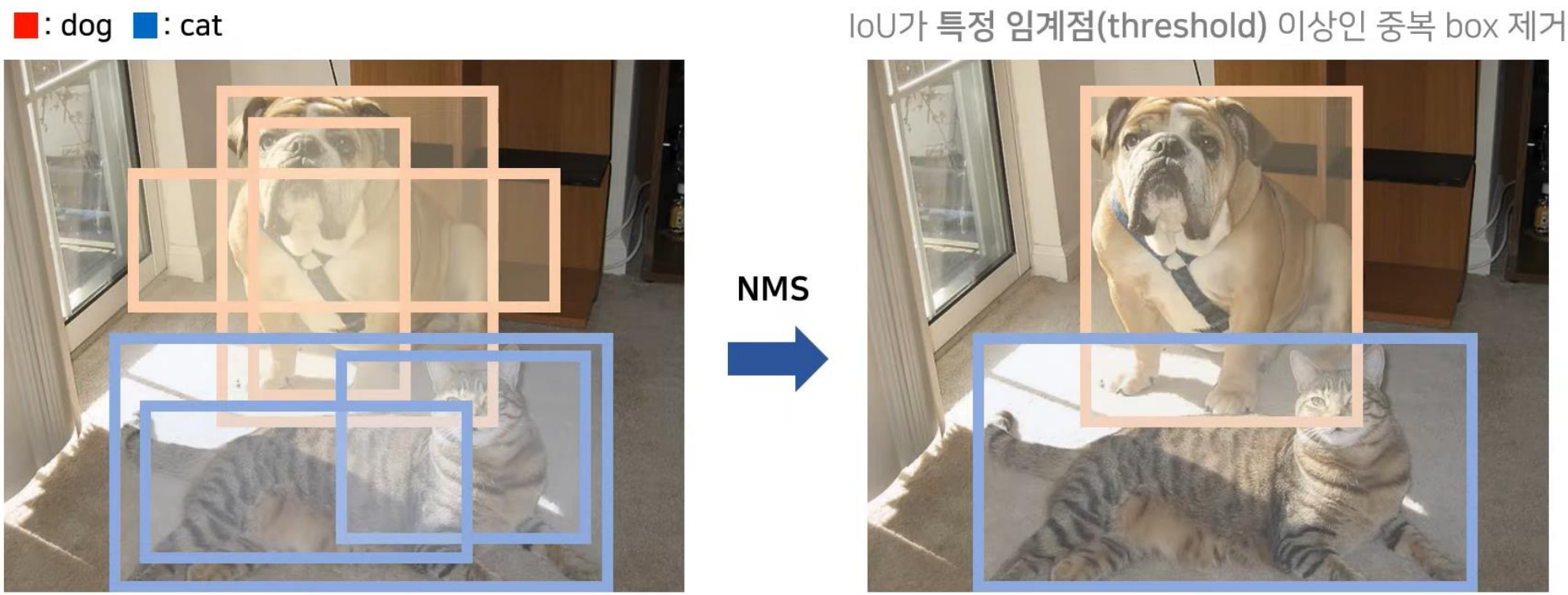
N_{cls}, N_{reg} = Weight Parameter

λ = Normalization Parameter (=10)

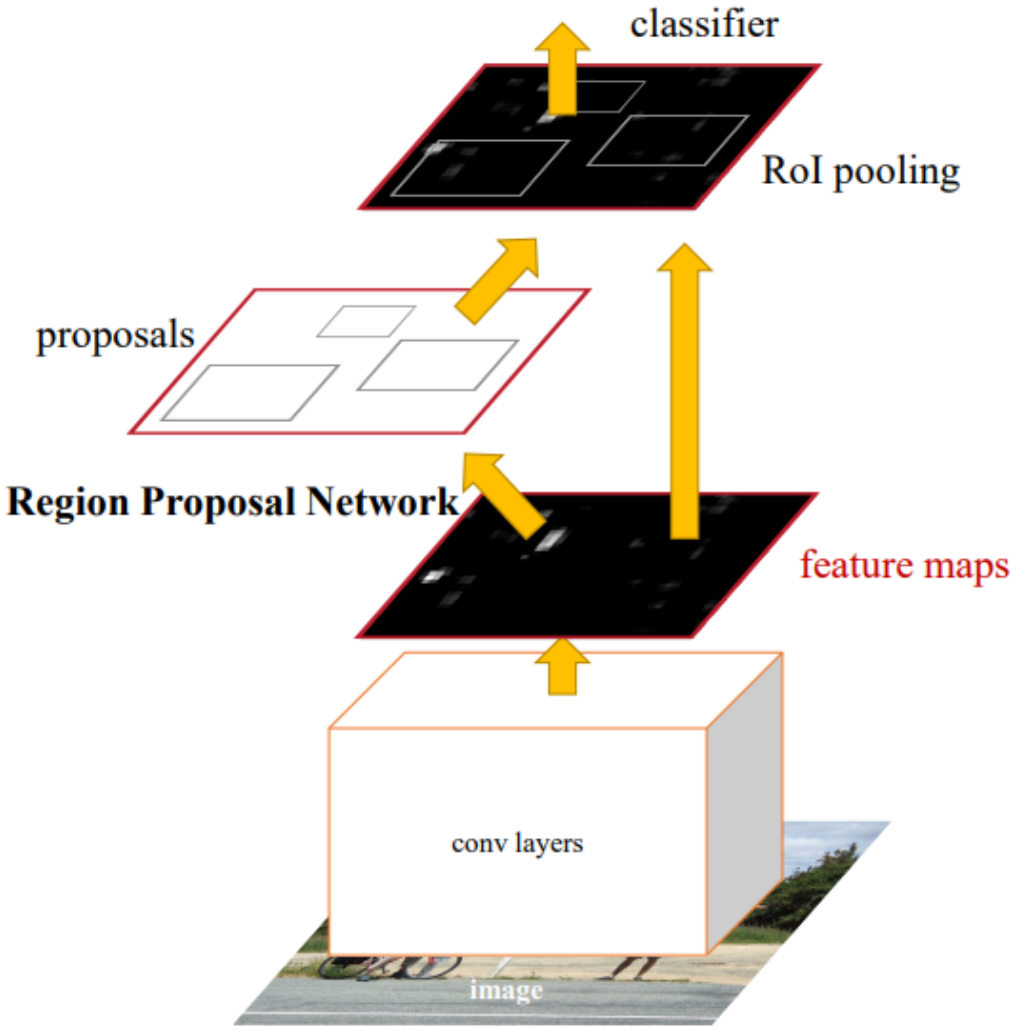
Faster R-CNN

Intersection Over Union(*IoU*) = 두 Bounding Box가 겹치는 비율

- **NMS 계산** : 같은 클래스(Class)끼리 IoU가 Threshold Value보다 이상일 때, 낮은 Confidence의 Bounding Box를 제거하는 방식
  (Non-Maximum Suppression)
  → 여러 개의 Bounding Box가 겹쳐 있는 경우에 하나로 합치는 방법!



Faster R-CNN


classifier
RoI pooling
proposals
Region Proposal Network
feature maps
conv layers
image

Experiments & Result

train-time region proposals		test-time region proposals		
method	# boxes	method	# proposals	mAP (%)
SS	2000	SS	2000	58.7
EB	2000	EB	2000	58.6
RPN+ZF, shared	2000	RPN+ZF, shared	300	**59.9**
ablation experiments follow below				
RPN+ZF, unshared	2000	RPN+ZF, unshared	300	58.7
SS	2000	RPN+ZF	100	55.1
SS	2000	RPN+ZF	300	56.8
SS	2000	RPN+ZF	1000	56.3
SS	2000	RPN+ZF (no NMS)	6000	55.2
SS	2000	RPN+ZF (no *cls*)	100	44.6
SS	2000	RPN+ZF (no *cls*)	300	51.4
SS	2000	RPN+ZF (no *cls*)	1000	55.8
SS	2000	RPN+ZF (no *reg*)	300	52.1
SS	2000	RPN+ZF (no *reg*)	1000	51.3
SS	2000	RPN+VGG	300	59.2

Experiments & Result

method	# proposals	data	mAP (%)
SS	2000	07	66.9†
SS	2000	07+12	70.0
RPN+VGG, unshared	300	07	68.5
RPN+VGG, shared	300	07	69.9
RPN+VGG, shared	300	07+12	**73.2**
RPN+VGG, shared	300	COCO+07+12	**78.8**

model	system	conv	proposal	region-wise	total	rate
VGG	SS + Fast R-CNN	146	1510	174	1830	0.5 fps
VGG	RPN + Fast R-CNN	141	**10**	47	**198**	**5 fps**
ZF	RPN + Fast R-CNN	31	**3**	25	**59**	**17 fps**

Experiments & Result

method	# box	data	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
SS	2000	12	65.7	80.3	74.7	66.9	46.9	37.7	73.9	68.6	87.7	41.7	71.1	51.1	86.0	77.8	79.8	69.8	32.1	65.5	63.8	76.4	61.7
SS	2000	07++12	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	89.3	44.2	73.0	55.0	**87.5**	80.5	80.8	72.0	35.1	68.3	**65.7**	80.4	64.2
RPN	300	12	67.0	82.3	76.4	71.0	48.4	45.2	72.1	72.3	87.3	42.2	73.7	50.0	86.8	78.7	78.4	77.4	34.5	70.1	57.1	77.1	58.9
RPN	300	07++12	70.4	84.9	79.8	74.3	53.9	49.8	77.5	75.9	88.5	45.6	77.1	55.3	86.9	81.7	80.9	79.6	40.1	72.6	60.9	81.2	61.5
RPN	300	COCO+07++12	75.9	87.4	83.6	76.8	62.9	59.6	81.9	82.0	91.3	54.9	82.6	59.0	89.0	85.5	84.7	84.1	52.2	78.9	65.5	85.4	70.2

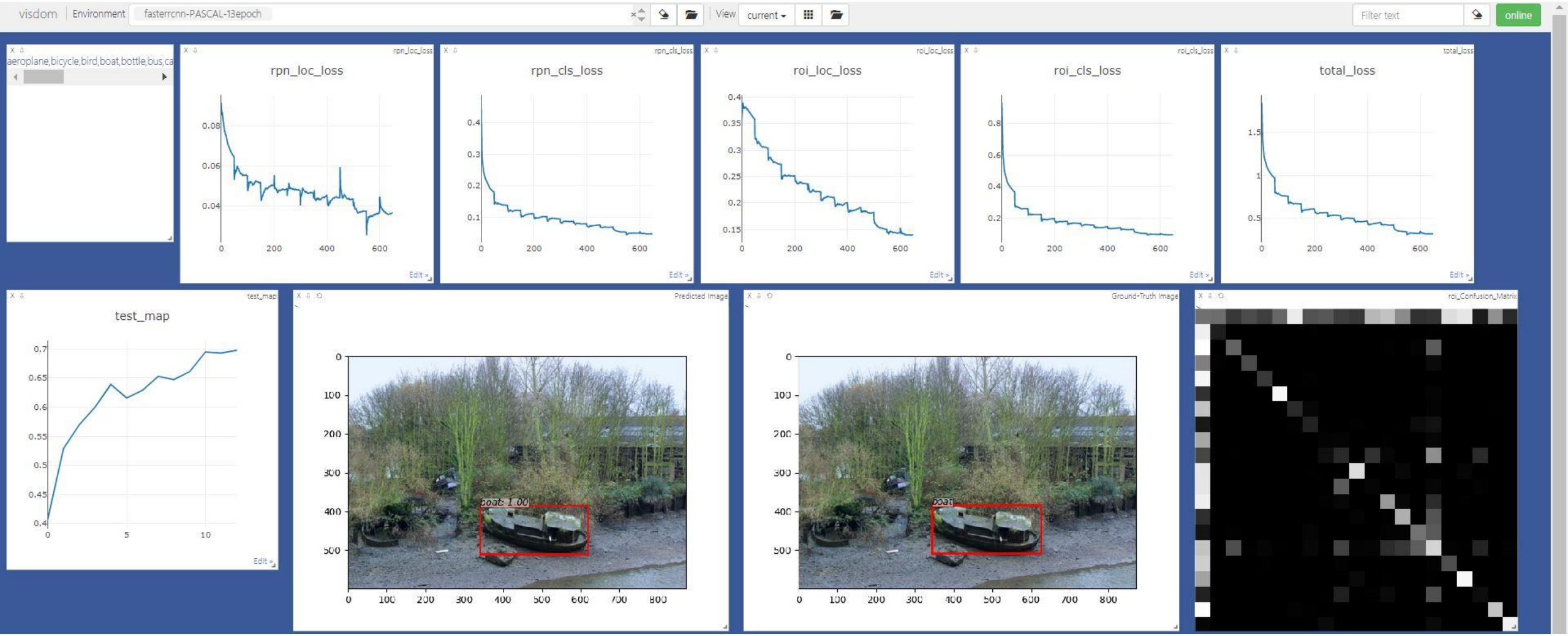
Experiments & Result

# proposals	50	100	150	200	300	500	1000
mAP (%)	66.3	68.9	69.5	69.8	69.9	69.8	69.8

λ	0.1	1	10	100
mAP (%)	67.2	68.9	69.9	69.1

settings	anchor scales	aspect ratios	mAP (%)
1 scale, 1 ratio	128^2	1:1	65.8
	256^2	1:1	66.7
1 scale, 3 ratios	128^2	{2:1, 1:1, 1:2}	68.8
	256^2	{2:1, 1:1, 1:2}	67.9
3 scales, 1 ratios	$\{128^2, 256^2, 512^2\}$	1:1	**69.8**
3 scales, 3 ratios	$\{128^2, 256^2, 512^2\}$	{2:1, 1:1, 1:2}	**69.9**

Experiments & Result



References

[Paper]

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

[Lecture]

https://www.youtube.com/watch?v=46SjJbUcO-c

https://www.youtube.com/watch?v=kcPAGIgBGRs&ab_channel=JinWonLee

https://www.youtube.com/watch?v=HmJWvwIpW5g&ab_channel=%EB%94%A5%EB%9F%AC%EB%8B%9D%EB%85%BC%EB%AC%B8%EC%9D%BD%EA%B8%B0%EB%AA%A8%EC%9E%84

[Blog]

https://blog.naver.com/kimsamuel351/222740515349

https://blog.naver.com/koreadeep/222661706323

https://foxtrotin.tistory.com/434

Q & A

Thank you ☺