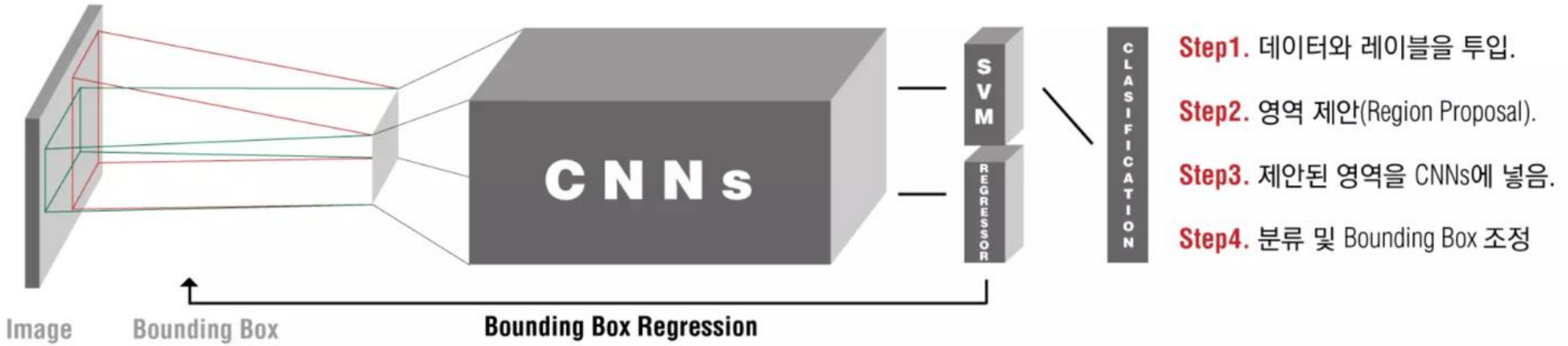
R-CNN의 정의와 구조
Definition of Convolution Neural Networks

1. R-CNN의 정의

Regions with Convolutional Neuron Networks의 약자로, 영역을 설정하고 CNNs을 활용하여 물체 인식(Object Detection)을 수행하는 신경망이다.

2. R-CNN의 구조와 절차



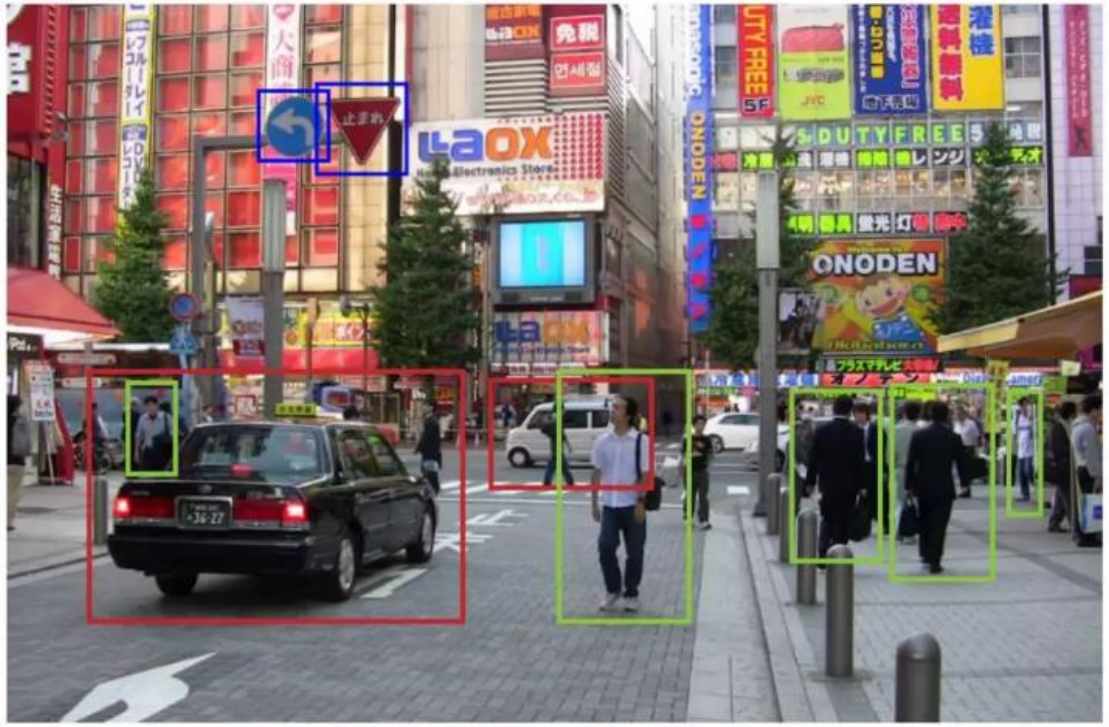
Step1. 데이터와 레이블을 투입.

Step2. 영역 제안(Region Proposal).

Step3. 제안된 영역을 CNNs에 넣음.

Step4. 분류 및 Bounding Box 조정

데이터와 레이블을 투입
Input the Images and Label



물체 인식(Object Detection)에서의 데이터와 레이블

; 데이터- 이미지(Image) // 레이블 - 정답 테두리상자(Bounding Box)

물체를 포함하는 영역을 사각형으로 가시화한 것

테두리 상자(BB)의 학습 과정

선택적 탐색(Selective Search)를 통해 임의의 BB를 설정

임의의 BB와 사전에 준비한 정답 BB의 IOU를 계산

IOU; Intersection Over Union

IOU가 특정 값 이상이 되도록 임의 영역을 조정

영역 제안; 선택적 탐색
Region Proposal by Selective Search

1. 선택적 탐색이란?

전부를 탐색하는 완전 탐색(Complete Search)과는 달리 특정 기준에 따라 탐색을 실시하는 것으로, 여기서는 상향식(Bottom-Up)의 탐색방법 중 하나인 계층적 그룹 알고리즘(Hierachical Grouping Algorithm)등이 사용되었다.

2. R-CNN에서의 선택적 탐색의 절차

선택적 탐색은 초기의 작은 크기의 세분화 영역을 설정하고, 이를 계층적 그룹 알고리즘을 사용하여 병합하고, 이를 바탕으로 영역을 제안하는 단계로 진행된다.

작은 크기의 초기 영역을 설정

Felzenszwalb&Huttenlocher - Segmentation

P. F. Felzenszwalb&D. P. Huttenlocher. Efficient GraphBased Image Segmentation. IJCV, 59:167–181, 2004

작은 영역을 큰 영역으로 병합

Hierachical Grouping Algorithm

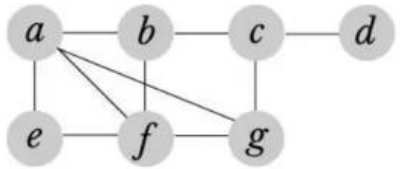
J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013

ROI(Regions of Interest) 설정
Region Proposals

작은 크기의 초기 영역을 설정
Initial Segmentation by Felzenszwalb&Huttenlocher

1. 그래프 이론 (Graph Theory)

그래프 이론은 수학에서 객체 간에 짝을 이루는 관계를 모델링하기 위해 사용되는 수학 구조인 그래프에 대한 연구이다. 그래프는 꼭짓점(vertex), 교점(node), 점(point)으로 구성되며, 이들은 선 또는 변(edge)으로 연결된다. (출처 – 위키백과)



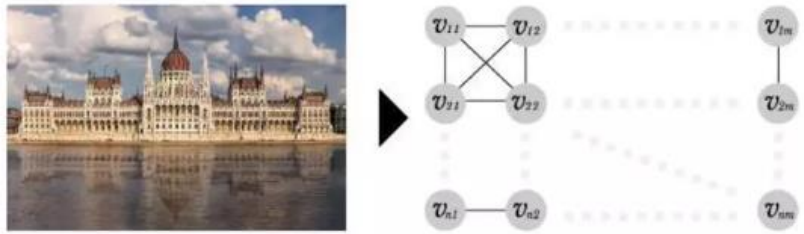
$$G = (V, E)$$

vertex: a, b, c, d, e, f, g

edge: (a,b), (a,e), (a,f), (a,g), (b,c), (b,f), (c,d), (c,g), (e,f), (f,g)

degree; d(x)=4 (x=a, f), d(y)=3 (y=b, c, e, g), d(d)=1

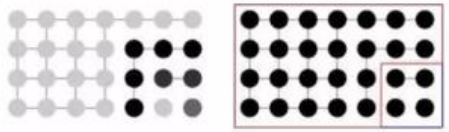
2. 이미지를 그래프로 표현



pixel → vertex $v_i \in V$ *connection(of pixel) → edge* $(v_i, v_j) \in E$

edge $e=(v_i, v_j)$ ~ *weight* $w(e)=w((v_i, v_j))=|I(p_i)-I(p_j)|$ *I(*)=Intensity of Pixel*

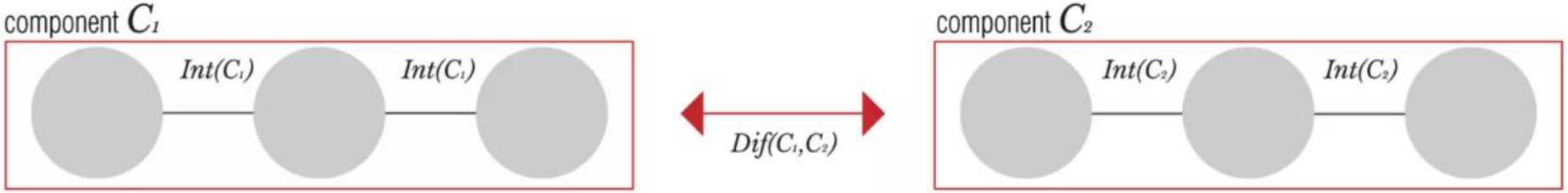
<Components> <Segmentation>



Each Component C (∈ S) Corresponds

Connected Component in G' = (V, E') E'⊆E

3. *Pairwise Region Comparison Predicate*



$$D(C_1,C_2) = \begin{cases} true & if\ Dif(C_1,C_2) > MInt(C_1,C_2) \\ false & \text{(영역 합침)}\ otherwise \end{cases}$$

$$MInt(C_1,C_2) = min\{Int(C_1)+\tau(C_1),\ Int(C_2)+\tau(C_2)\}$$

$$\tau(C) = k / |C|$$

$$Int(C) = \max_{e \in MST(C,E)} w(e)$$

MST;
Minimum Spanning Tree

$$Dif(C_1,C_2) = \min_{v_i \in C_1,\, v_j \in C_2,\, (v_i,v_j) \in E} w((v_i,v_j))$$

4. 전체 알고리즘

IMAGE *Graph G=(V, E)*

1. Sort E into $\pi = (o_1, \ldots, o_m)$ by non-decreasing edge weight.
2. Start with a segmentation S^0, where each vertex v_i is in own component.

for q in range(1, m+1):

Construct S^p given S^{p-1}

Let v_i and v_j denote the vertices connected by the q-th edge in ordering i.e. $o_q = (v_i, v_j)$

Let C_i^{q-1} be the component of S^{q-1} containing v_j, C_j^{q-1} the component containing v_j

$$D(C_i^{q-1}, C_j^{q-1}) = \begin{cases} S^q = S^{q-1} & \text{if } Dif(C_i^{q-1}, C_j^{q-1}) > MInt(C_i^{q-1}, C_j^{q-1}) \\ S^q \text{ is obtained by merging } C_i^{q-1}, C_j^{q-1} & \text{otherwise} \end{cases}$$

Return $S = S^m$

Segmented IMAGE *Segmentation of V*

작은 영역을 큰 영역으로 병합; 계층적 병합 알고리즘

Merging the Segmentations; Hierachical Grouping Algorithm

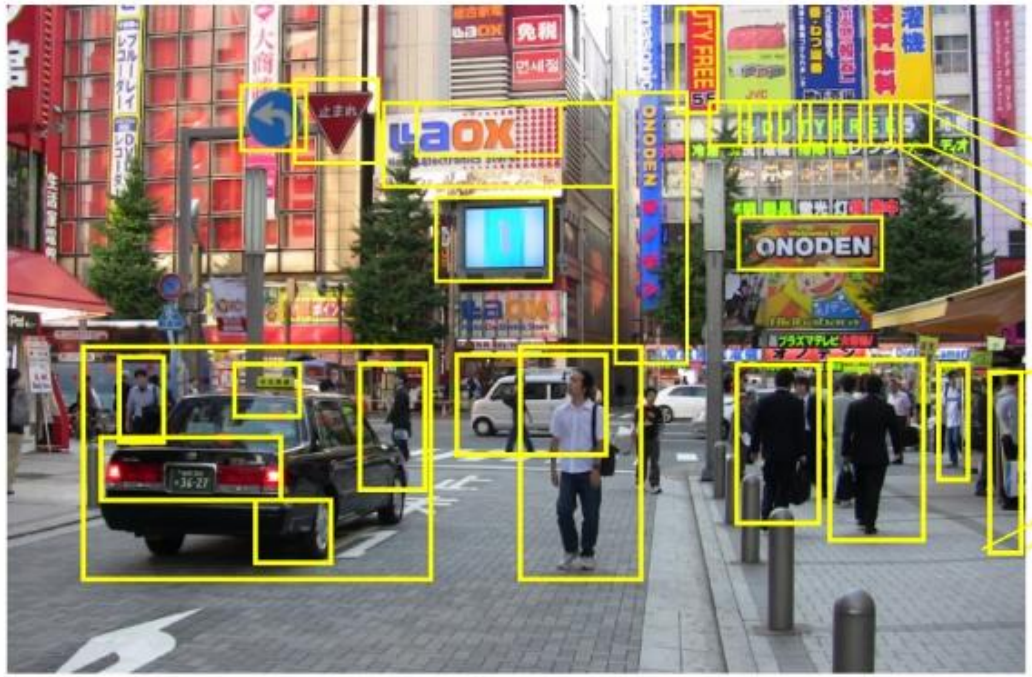
(COLOUR) IMAGE

1. *Obtain initial regions* $R=\{r_1, \cdots, r_n\}$ *by Felzenszwalb & Huttenlocher's Segmentation(Previous Method).*
2. *Initialize similarity set* $S = \phi$
3. *foreach Neighbouring region pair* (r_i, r_j) *do*
   Calculate similarity $s(r_i, r_j)$, $S=S \cup s(r_i, r_j)$
4. *while* $S \neq \phi$ *do*
   Get highest similarity $s(r_i, r_j) = max(S)$, *Merge corresponding regions* $r_t = r_i \cup r_j$
   Remove similarities regarding r_i : $S=S \setminus s(r_i, r_*)$, *Remove similarities regarding* r_j : $S=S \setminus s(r_*, r_j)$
   $S=S \cup S_i$, $R=R \cup r_i$
5. *Extract object location boxes L from all regions in R*

Set of object location hypothesis L (Region Proposal)

제안된 영역을 CNNs의 입력값으로

Region Proposals as A Inputs of CNNs



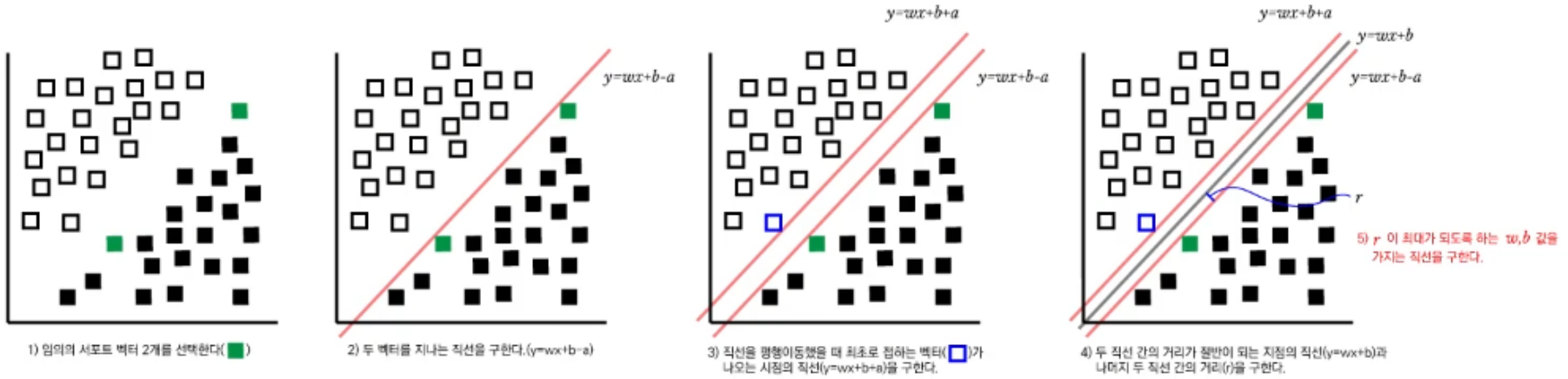
서로 다른 크기의 ROI를 CNNs의 **정해진 크기**로 맞추어서
각각을 입력값으로 입력

WRAPPING

분류 및 테두리상자 조정
Classification and Bounding Box Regression

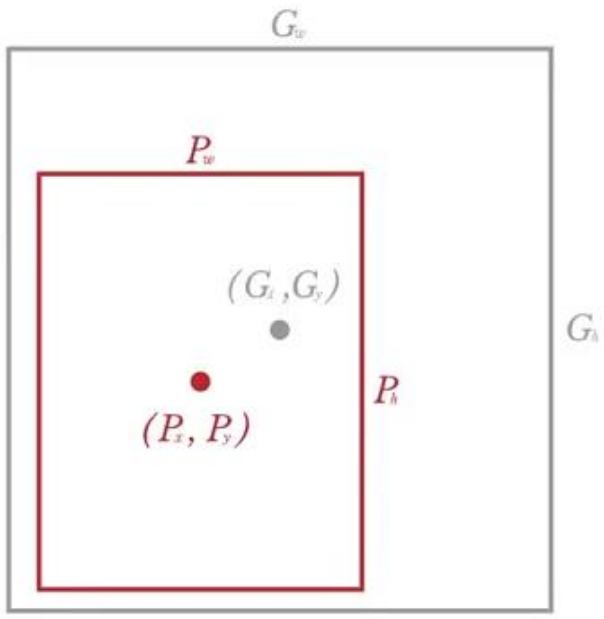
1. 서포트 벡터 머신(SVM)을 통한 분류

SVM은 기계 학습의 분야 중 하나로 패턴 인식, 자료 분석 등을 위한 지도 학습 모델이며, 주로 분류와 회귀 분석을 위해 사용한다.



각각의 분류 클래스에 대해서 SVM을 적용한 Matrix를 기반으로 분류 작업을 시행 (해당 클래스 ~ O / X)

2. 테두리상자 조정 (Bounding-Box Regression)



P_* ;Bounding Box $\hat{G}_*$;Optimized Bonding Box

G_* ;Ground Truth $d_*(\cdot)$;Transformation Function

$$P^i = (P_x^i, P_y^i, P_w^i, P_h^i) \quad \blacktriangleright \quad G = (G_x, G_y, G_w, G_h)$$

$$\hat{G}_x = P_w d_x(P) + P_x \qquad \hat{G}_y = P_h d_y(P) + P_y$$

$$\hat{G}_w = P_w \exp(d_w(P)) \qquad \hat{G}_h = P_h \exp(d_h(P))$$

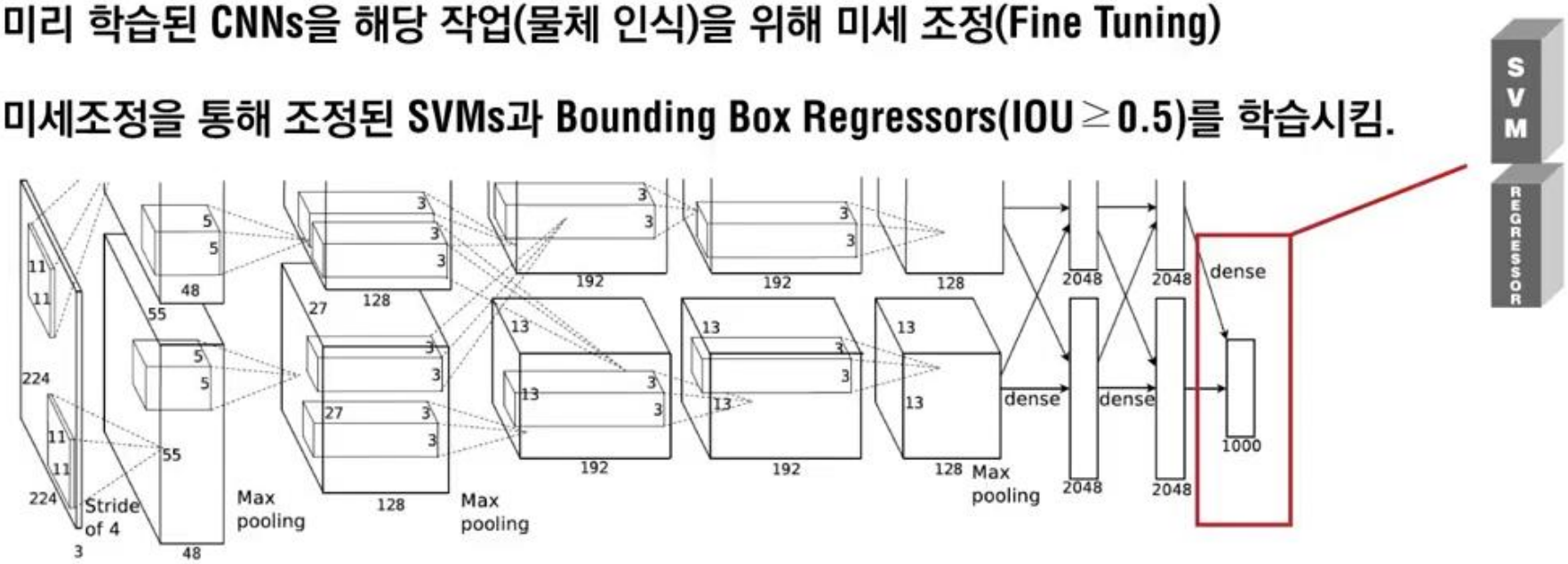
$$w_* = \underset{\hat{w}_*}{argmin} \sum_i^N (t^i - \hat{w}_*^T \Phi_5(P^i))^2 + \lambda \|\hat{w}_*\|^2$$

$$t_x = (G_w - P_w)/P_w \qquad t_y = (G_h - P_h)/P_h$$

$$t_w = \log(G_w/P_w) \qquad t_h = \log(G_h/P_h)$$

RCNN 학습 과정
Process of Training RCNN

1. ImageNet의 데이터셋을 바탕으로 CNNs(논문에서는 AlexNet)을 미리 학습
2. 미리 학습된 CNNs을 해당 작업(물체 인식)을 위해 미세 조정(Fine Tuning)
3. 미세조정을 통해 조정된 SVMs과 Bounding Box Regressors(IOU $\geq$ 0.5)를 학습시킴.



R-CNN 한계점

- 합성곱 신경망의 입력을 위한 고정된 크기를 위해 warping/crop을 사용해야하며 그 과정에서 이미지 정보 손실이 일어난다.
- 2000개의 영역마다 CNN을 적용해야 하기에 학습 시간이 오래 걸린다.
- 학습이 여러 단계로 이루어지며 이로 인해 긴 학습 시간과 대용량 저장 공간이 요구된다.
- Object Detection의 속도 자체도 느리다.