

7주차 복습과제 (이론) - 문제

1. You are building a 3-class object classification and localization algorithm. The classes are: pedestrian (c=1), car (c=2), motorcycle (c=3). What would be the label for the following image? Recall $y=[pc, bx, by, bh, bw, c1, c2, c3]$



답) $y = [1, 0.3, 0.7, 0.3, 0.3, 0, 1, 0]$

2. Continuing from the previous problem, what should y be for the image below? Remember that “?” means “don’t care”, which means that the neural network loss function won’t care what the neural network gives for that component of the output. As before, $y=[pc, bx, by, bh, bw, c1, c2, c3]$.



- 1) $y = [?, ?, ?, ?, ?, ?, ?, ?]$
- 2) $y = [1, ?, ?, ?, ?, 0, 0, 0]$
- 3) $y = [1, ?, ?, ?, ?, ?, ?, ?]$
- 4) $y = [0, ?, ?, ?, ?, ?, ?, ?]$
- 5) $y = [0, ?, ?, ?, ?, 0, 0, 0]$

답) 4번

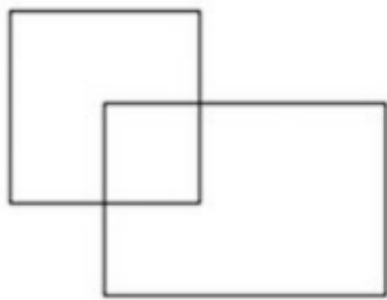
3. If you build a neural network that inputs a picture of a person's face and outputs N landmarks on the face (assume the input image always contains exactly one face), how many output units will the network have?

답) $2N$

4. When training one of the object detection systems described in lecture, you need a training set that contains many pictures of the object(s) you wish to detect. However, bounding boxes do not need to be provided in the training set, since the algorithm can learn to detect the objects by itself. (True or False)

답) True

5. What is the IoU between these two boxes? The upper-left box is 2×2 , and the lower-right box is 2×3 . The overlapping region is 1×1 .



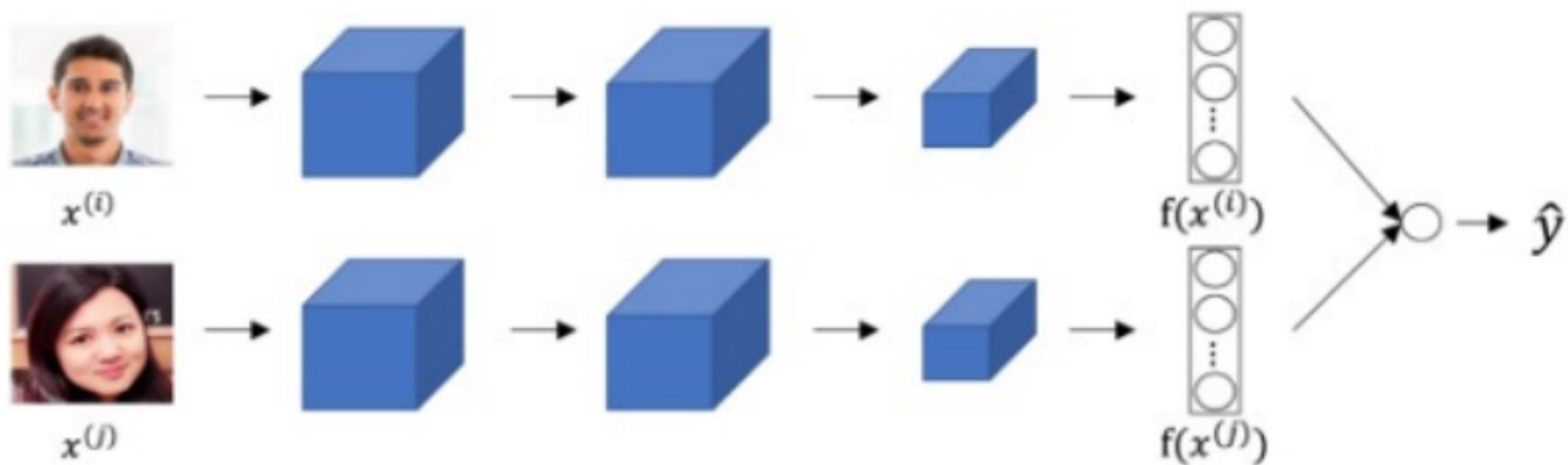
답) $1/9$

6. Why do we learn a function $d(img1, img2)$ for face verification? - (주관식 작성)

답)

- 적은 데이터셋 처리하기 위해
- 비지도 방식 차용하기 위해
- 삼 네트워크 통한 원샷 퓨샷 러닝 활용하기 위해

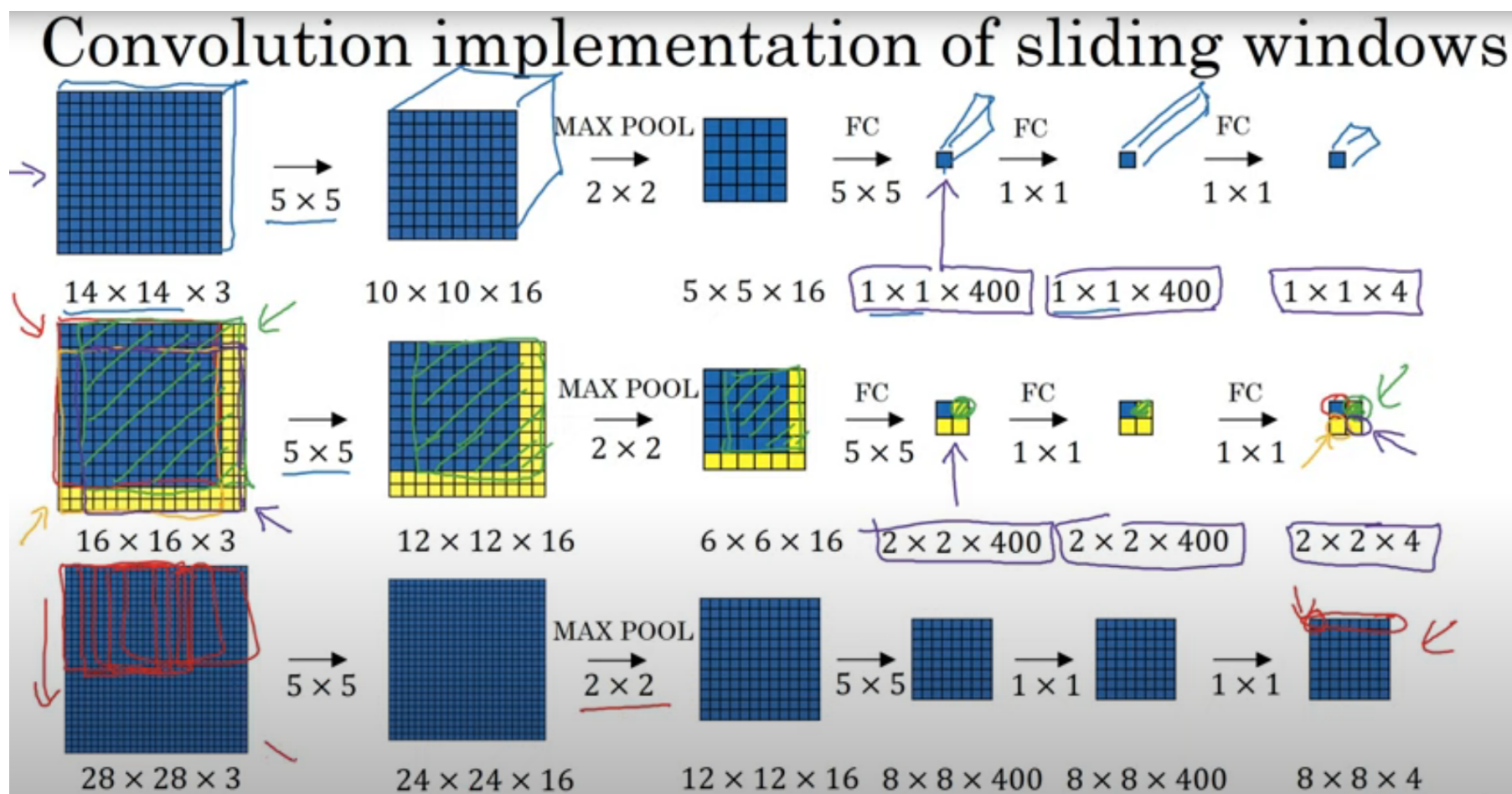
7. Consider the following Siamese network architecture:



The upper and lower neural networks have different input images and the different parameters.

답) False

8. C4W3L04의 강의에서는 sliding window에 convolution을 적용하여 연산 속도를 빠르게 하는 방법을 설명하고 있습니다. 아래의 강의 슬라이드를 자신이 이해한대로 설명하는 글을 작성해주세요.



답) Sliding Window Approach는 Convolution의 컨셉을 내재하고 있습니다. 합성곱을 통해 Sliding Window 합성곱 크기와 같은 Reception Field의 특징을 Feature map 형태로 추출하고, 이를 FC layer에 통과시키면서 정답 레이블을 담고 있는 결과 벡터를 추출합니다. 이미지의 크기가 더 커진 경우에도 행렬곱과 병렬 처리에 최적화된 현대 하드웨어와 여러 연산이 중첩되는 Convolution의 특징을 통해 computational cost를 줄일 수 있습니다. Semantic하게 생각해보면, 맨 마지막 그림의 output은 총 64개 슬라이딩 윈도우의 detection 결과를 원한 인코딩 형태 (혹은 확률값으로)로 표현합니다.