1. What is the most ACCURATE and PRECISE definition of the camera obscura?

a. A passive exteroceptive sensor

b. A mathematical model that describes relationship between world coordinates of a point and its projection onto the image plane

c. A box or room with a pinhole aperture in front of an imaging surface

d. A tiny hole in a barrier through which light travels

e. None of the above

answer: c
reason: camera obscura는 그림을 사실적으로 그리기 위해 작은 구멍을 뚫은 상자이다.

2. Which of the following statements are TRUE? Select all that apply.

a. Camera intrinsic parameters include a translation vector.

b. Camera extrinsic parameters include a rotation matrix.

c. Camera extrinsic parameters define the transformations from 3D world coordinates to 3D camera coordinates

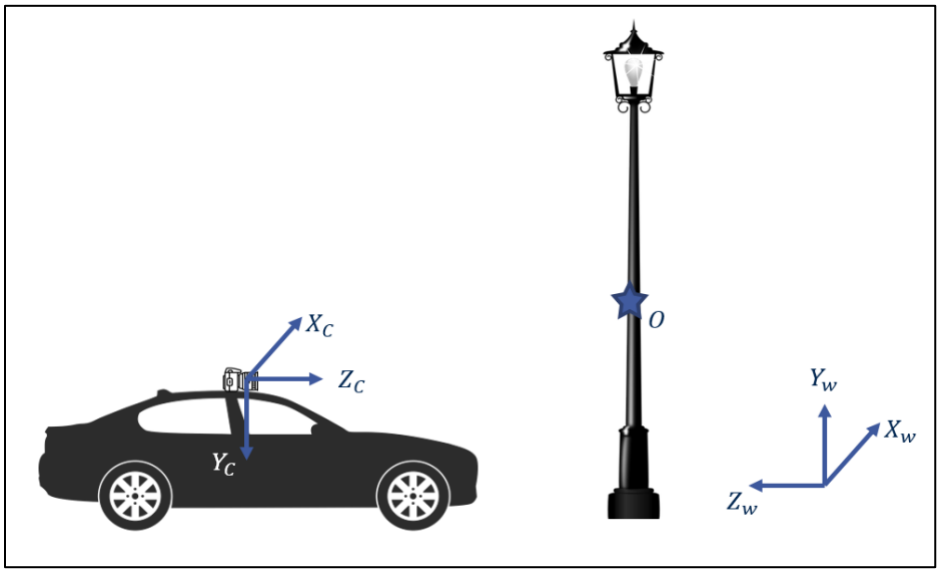
d. Camera extrinsic parameters include the focal length.

e. Camera intrinsic parameters define the transformations from 3D camera coordinates to 3D world coordinates

answer: b
reason: Camera 외부 변수는 높이, 방향, 회전을 포함하며 회전 및 평행 이동 변환으로 표현된다.

3. Imagine a situation in which a camera mounted on a car sees a point O on a signpost. The location of this point in world coordinate system is $O = \begin{bmatrix} -0.5 \\ 1.5 \\ 9 \end{bmatrix}$ meters. The location of the center of the world coordinate system relatively to the camera optical center in camera frame coordinates is defined by the translation vector $t = \begin{bmatrix} 1 \\ 2 \\ 10 \end{bmatrix}$ meters and 180° rotation around the X_C axis.

The camera intrinsic parameter matrix is $K = \begin{bmatrix} 640 & 0 & 640 \\ 0 & 380 & 380 \\ 0 & 0 & 1 \end{bmatrix}$ and the image resolution is 1280 X 960 pixels. You can check out the image below to help you visualize the problem. **What is the position of this point in the camera coordinate system?** Please write your answer as a string with three comma-separated numbers without spaces (x and y and z), e.g. **"1,2,3"**



answer: 0.5,0.5,1

reason: [1001;0-102;00-110]*[-0.5 1.5 9 1]T= [0.5 0.5 1]T

4. Variation of prob., 3

Based on the problem presented in the above question, what is the pixel location of the 2D projection of the point O on the image plane?

Please write your answer as a string with two comma-separated numbers without spaces (u and v), e.g. "100,100"

answer: 960,570
reason: K*[0.5 0.5 1]T

5. Why is camera calibration important in the self-driving car domain? Select all that apply.

a. Computed camera parameters can be used to determine the camera model

b. Computed camera parameters can be used to determine the camera location relative to the scene

c. Computed camera parameters can be used to correct for lens distortion

d. Camera calibration estimates the parameters of the lens and image sensor of a camera

e. Computed camera parameters can be used to measure the size of a 2D object in 3D world units

answer: b, c, d, e
reason: Camera calibration으로 camera model을 알 수는 없으며 camera location 을 파하기 위해world coordination에서 image coordination으로 변환하는 과정에서 lens distortion을 정정하여 영점을 맞출 수 있다.

6. Recall the camera calibration problem formulation, which has the following mathematical representation.

What methods from linear algebra can we use for solving this problem? Select all that apply.

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_NX_N & -u_NY_N & -u_NX_N & -u_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & X_1 & -v_NX_N & -v_NY_N & -v_NY_N & -u_N \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \\ p_{34} \end{bmatrix} = 0$$

a. Gaussian Elimination

b. Method of Complements

c. Eigen Decomposition

d. Singular Value Decomposition

e. None of the above

answer: d
reason: 3D points를 2D로 projection할 때 homogeneous 선형 system을 설정하여 행렬을 factorization하는 singular value decomposition으로 해결한다.

7. Let's continue with the camera calibration formulation from the previous question. What are some DISADVANTAGES of this linear calibration model? Select all that apply.

a. The linear model mixes the intrinsic and extrinsic camera parameters

b. The linear model is computationally expensive

c. The linear model does not allow for the incorporation of parameter constraints

d. Does not model radial distortion and other complex phenomena

e. None of the above

answer: a, c, d
reason: Camera calibration으로 camera parameter를 직접적으로 구할 수 없고 초점거리 제한 조건을 허용하지 않고 복잡한 형태, 비선형적 모델을 구할 수 없다.

8. If the baseline between camera centers is known for a stereo rig, what limitation of monocular vision can be avoided?

a. Motion blur

b. Inability to measure depth to a point

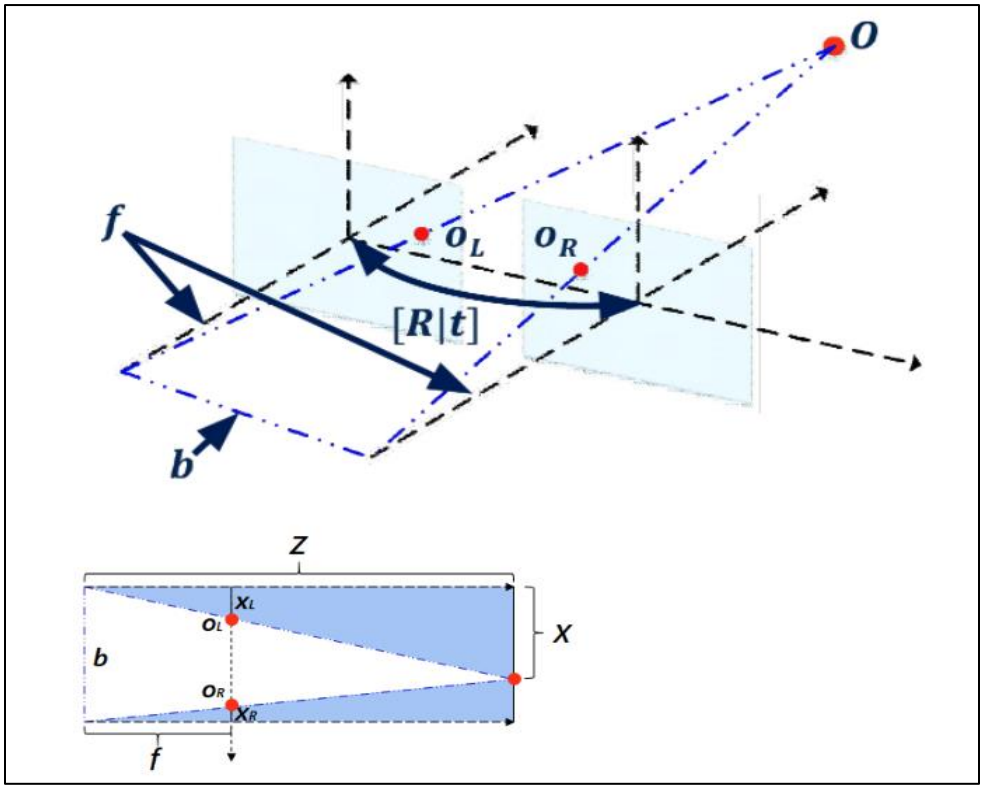
c. Over exposure in bright lighting conditions

d. Inability to measure angle to a point

e. None of the above

answer: b
reason: Stereo camera를 사용하면 단일 카메라를 사용하는 것과는 다르게 stereopsis 계산을 통해 두 사진의 영점을 맞추어 사물과 카메라 사이의 거리를 알 수 있다.

9. Consider a stereo camera setup in the figures below, similar to what you saw in the course slides. Which of the statement about this configuration are correct? Select all that apply.



a. $\frac{x_L}{f} = \frac{b - X}{Z}$

b. $\frac{Z}{X} = \frac{f}{x_R}$

c. $\frac{x_R}{f} = \frac{X - b}{Z}$

d. $\frac{Z}{X} = \frac{f}{X_L}$

e. $\frac{X_R}{f} = \frac{b - X}{Z}$

answer: d, e

reason: (d) Z : f = X : XL, (e) Z : f = b-X : XR

10. What parameters and computations are needed to perform depth calculations from disparity measurements? Select all that apply.

a. Need to compute/know the focal length

b. Need to find the extrinsic transformation between the stereo cameras and the world reference frame

c. No further information is needed. Depth can be computed directly from the pixel disparity.

d. Need to compute the baseline, and the x and y offsets for the image center

answer: a, d
reason: Disparity computing을 하기 위해서 crresponce를 위한 epipolar line을 긋고 optical axes가 평행할 때는 왼쪽 이미지의 pixel과 동일 선상 오른쪽 이미지의 pixel과 비교한다. 이 연산을 할 때 focal depth, center 사이의 거리를 알아야 한다.

11. A naive solution for the stereo correspondence problem is an exhaustive search, where we search the whole right image for a match to every pixel in the left image. Why is this a bad approach? Select all that apply

a. Radial and tangential distortion make it difficult to match corresponding pixels with this approach because it distorts each image differently

b. It is a good approach, and none of the concerns above is valid

c. This approach generally could not run in real time

d. An exhaustive search can be performed only when the left and right images are taken with the same camera model. Otherwise, the images are too different and cannot be compared

e. The naive approach results in a large number of incorrect matches due to similar pixels in different parts of the image

answer: c, e
reason: 왜곡이 생긴 경우(optical axes가 평행하지 않은 경우)에는 stereo rectification 단계를 거쳐 정정할 수 있다. 왼쪽 이미지 pixel을 동일 선상 오른쪽 이미지 pixel과 비교하며 minimum cost를 판별하는 것은 상당히 오랜 시간이 걸리며 정확한 계산에는 어려움이 있다.

12. What is the definition of an epipolar line for stereo cameras?

a. A straight line connecting the left and right camera centers in a stereo setup

b. A straight line connecting the optical center of a camera and a point in the scene

c. A straight line that passes through the center of the lens and the camera sensor

d. A line produced in one camera as a point in 3D space is moved along a single ray emanating from the other camera's optical center

e. None of the above

answer: d
reason: Epipolar line은 3차원의 한 점과 두 카메라 원점을 잇는 평면과 이미지 평면과의 교선이다.

13. Recall that the first basic stereo algorithm which you saw in this module has four necessary steps. These steps are given to you below. Your task is to put them into correct order

(A) Compute disparity

(B) Pick the pixel that has minimum cost

(C) Consider each pixel on the epipolar line in the left image

(D) Compare the chosen pixels from the left image to every pixel in the right image on the same epipolar line

a. (D), (C), (B), (A)

b. (C), (D), (A), (B)

c. (C), (D), (B), (A)

d. (C), (B), (D), (A)

answer: c
reason: Disparity computing은 왼쪽 이미지에서 epipolar line을 그어 같은 선 위의 오른쪽 이미지와 비교하고 저렴한 경우를 선택하여 계산하는 과정으로 진행된다.

14. Which of the below statements about correlation and convolution are correct? Select all that apply.

a. A convolution kernel is a 180 degrees rotated cross-correlation kernel

b. The order of multiplication of convolution kernels does not matter

c. Template matching can only be performed with a convolution filter and not with a cross-correlation filter

d. A convolution filter and a cross-correlation filter are the same if the kernel is symmetric

e. Cross-correlation is associative

answer: a, b, d
reason: Convolutiondms cross-correlation과 달리 결합법칙 적용이 가능하여 곱하는 순서가 정해져있지 않으며 cross correlation 연산 결과가 filp된 형태이다.

15. Which of these 3X3 image filters is a Gaussian filter?

a. $\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

b. $\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$

c. $\frac{1}{10}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

d. $\frac{1}{9}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$

e. None of above

answer: b

reason: Gaussian filter는 중심으로 갈수록 가중치(행렬 원소의 값)가 높아진다.

16. A feedforward neural network has an input layer, 5 hidden layers and an output layer. What is the depth of this neural network?

answer: 6
reason: N layer Feed Forward Neural Networks를 수행할 때 f1부터 f(N-1)까지를 hidden layer라고 한다.

17. During training, the training data specifies the exact form of the hidden layers of a neural network.

a. True

b. False

answer: b
reason: Training data는 training 과정을 거친 후에 결정된다.

18. Implement the ReLU activation function using numpy by replacing None in the code bellow.

```
import numpy as np

def ReLU(x):

  y = None

  return y
```

answer: y = max(0,x)
reason: ReLU activation function은 두 벡터가 유사할 때만 활성화되므로 입력된 값이 음수이면 0, 그렇지 않으면 입력된 값 그대로 출력된다.

19. The main building blocks of a machine learning system are: (Check all that apply.)

a. Hidden layers

b. A Model

c. A loss function

d. An Optimization Procedure

e. Output Layers

answer: b, c, d
reason: Machine learning은 한 model에 대한 예측과 실제 값의 차이를 줄이는 loss function, 그의 출력값인 output layers로 구성된다.

20. Which output unit/loss function pair is usually used for regression tasks that use neural networks?

a. Sigmoid output units with Mean Squared Error Loss

b. Linear output units with Mean Squared Error Loss

c. Linear output units with Cross-Entropy Loss

d. Softmax output units with Cross-Entropy Loss

answer: b
reason: Regression output unit은 비선형이 아니며 실제값과 예측값의 차의 제곱의 총합으로 구한다.

21. The softmax output layer with cross-entropy loss is used to model the mean of a Gaussian distribution.

a. True

b. False

answer: b
reason: Gaussian modeling에는 corss correlation 연산이 적용된다.

22. Which of the following might be used as a stopping condition for gradient descent. (Check all that apply.)

a. The value of the training loss

b. The magnitude of change in loss function value

c. The number of iterations or epochs

d. The magnitude of the change in parameter values

answer: b, c, d

reason: Gradient descent는 정해진 반복 횟수, threshold 미만의 parameter 변화, threshold 미만의 loss function 값의 변화 만족 시 종료된다.

23. How are neural network bias parameters usually initialized at the beginning of training?

a. Initialized to -1.

b. Initialized to samples from a standard normal distribution.

c. Initialized to samples from a standard uniform distribution.

d. Initialized to 0.

answer: d
reason: Biases는 초기에 0으로 설정하고 weight는 random 값으로 설정한다.

24. Using all samples to estimate the gradient of the loss function with respect to the parameter results in less than linear return in accuracy of this estimate.

a. True

b. False

answer: a
reason: Gradient descent의 결과로 선형 반환보다 작은 정확도가 나타난다.

25. You are working on a self-driving car project and want to train a neural network to perform traffic sign classification. You collect images with corresponding traffic sign labels, and want to determine the number of frames you will use for training. Given that you have around one million images with labels, what training/validation/testing data split would you use?

a. 100% training, 0% validation, 0% testing.

b. 20% training, 40% validation, 40% testing.

c. 60% training, 20% validation, 20% testing.

d. 96% training, 2% validation, 2% testing.

answer: d
reason: Training data는 많을수록 학습에 도움이 된다. 10000개의 data가 있으면 그 중 2%라 하더라도 2000개의 validation data, testing data가 확보되므로 d가 가장 적절하다.

26. You finish training your traffic sign classifier, and want to evaluate its performance. You compute the classification accuracy on the training, validation, and testing data splits and get the following results:

Data Split	Training	Validation	Testing
Accuracy	70%	68%	67%

You know that a human has an accuracy of around 98% on the traffic sign classification task. What are things you might try to achieve human level performance? (Check all that apply.)

a. Add regularization to your neural network.

b. Train your neural network longer.

c. Add more layers to your neural network.

d. Collect more training data.

answer: b, c
reason: 정확성이 70% 이하라는 건 학습이 부족하여 방향성을 잡지 못했기 때문이다. 따라서 더 오래 train해야하고 더 많은 layer를 통해 underfitting을 해결해야한다.

27. When a neural network overfits the training data, the generalization gap is usually very small.

a. True

b. False

answer: b
reason: Overfit의 경우 특정 data에 지나치게 특화되어 일반화하기에 어려움이 크다.

28. Which of the following strategies are used for regularization in neural networks? (Check all that apply.)

a. Norm Penalties

b. Training the neural network longer

c. Early Stopping

d. Increasing the number of parameters in the neural network architecture

e. Dropout

answer: a, c, e
reason: Regularization의 세 가지 방법으로는 parameter norm penalties, dropout, early stopping이 있다.

29. Dropout significantly limit the type of neural network models that can be used, and hence is usually used for specific architectures.

a. True

b. False

answer: b
reason: Dropout은 일부 뉴런을 제거하여 동일한 output layer을 얻어내는 과정이므로 특정 상황에만 적용되는 것이 아니다.

30. The name convolutional neural networks comes from the fact that these neural networks use a convolution operation instead of general matrix multiplication.

a. True

b. False

answer: b
reason: 분류 단계에서 covolutional layer가 적용되기 때문에 convolutional neural network이다. 또한 CNN에서는 filp이 없는 cross correlation 연산이 적용된다.

31. The input to a pooling layer has a width, height and depth of 224x224x3 respectively. The pooling layer has the following properties:

- Kernel shape: 2x2
- Stride: 2

What is the width of the output of this pooling layer?

answer: 112
reason: Wout = (Win-n) / s + 1 이므로 (224-2) / 2 + 1 = 112 이다.

32. Using convolutions might reduce overfitting, as the number of parameters in convolutional layers is less than the number of parameters in fully connected layers.

a. True

b. False

answer: a
reason: Fully connected layers는 parameter가 지나치게 많아 비효율적이며 convolutional layers는 그에 비해 parameter가 적어 overfitting 발생 가능성이 낮다.

33. The object detection problem is defined as the locating objects in the scene, as well as classifying the objects' category.

a. True

b. False

answer: a
reason: Object detection problem은 location과 label(classification)으로 구성된다.

34. The problem of object detection is non-trivial. Which of the following statements describe reasons for the difficulty in performing object detection? (Check all that apply.)

a. Extent of objects is not fully observed.

b. Cameras are not reliable to perform detection in outdoor environments.

c. The objects that are usually of interest to detect are highly variable in shape and color.

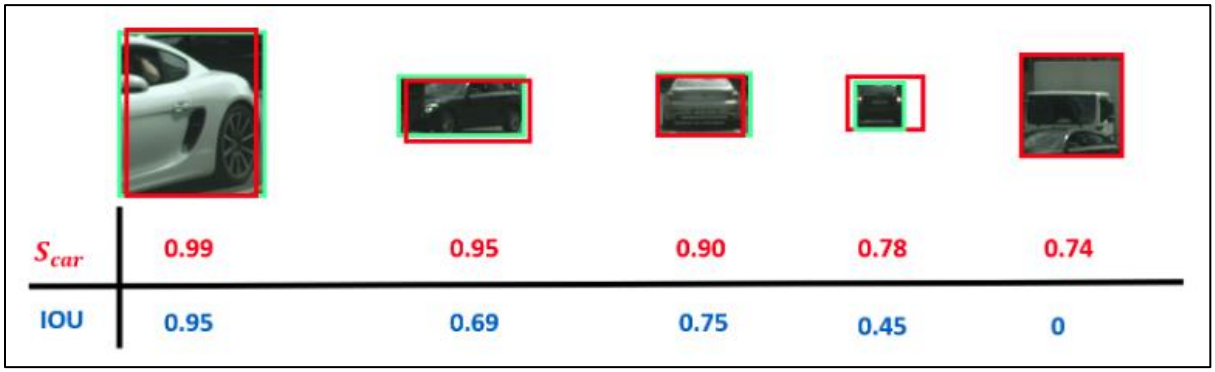
d. Scene illumination is highly variable on road scenes.

e. Object size gets smaller as objects move farther away in a road scene.

answer: a, d, e
reason: Object detection은 개체가 이미지 경계를 벗어나는 것(truncation), 지나치게 밝거나 어두운 밝기, 멀리 있는 개체가 작게 보이는 것(scale) 때문에 매우 복잡한 과정이다.

35. You are a self-driving car perception engineer developing an object detector for your self-driving car. You know that for your object detector to be reliable enough to deploy on a self-driving car, it should have a minimum precision of 0.99 and a minimum recall of 0.9. The precision and recall are to be computed at a score threshold of 0.9 and at an IOU threshold of 0.7.

You compute the IOU of your detector on a frame with ground truth to find out the following:



Assuming that the single frame shown above is sufficient to characterize the performance of the object detector, is your system reliable to be used on a self-driving car?

a. Yes

b. No

answer: b
reason: 위의 사진 순서대로 TP, FN, TP, FN, TN이다. 따라서 precision은 2/2=1이며 recall은 2/4=0.5이므로 threshold에 만족하지 않는다.

36. The width and height of the output of a convolutional feature extractor are usually an order of magnitude higher than those of its input.

a. Yes

b. No

answer: b
reason: Feature extractor의 출력은 입력 이미지의 출력에 비해 width, height가 작고 depth가 크다.

37. The input to a convolutional layer has a width, height and depth of 224x224x3 respectively. The convolutional layer has the following properties:

- Kernel shape: 3x3x256
- Stride: 2
- Padding: 3

What is the depth of the output of this convolutional layer?

answer: 256
reason: Depth(out)은 filter의 개수와 같다. Kernel shape에서 filter의 개수는 256개이므로 depth of the output은 256이다.

38. When designing convolutional architectures for object detection, max pooling layers are usually placed in which of the following building blocks:

a. Prior anchor boxes

b. Loss function

c. Output fully connected layers

d. Convolutional feature extractor

answer: d
reason: Feature extract 단계에서 convolutional layer, pooling layer 중 한 가지가 사용된다.

39. What type of output layer is most commonly used in the regression head of a convolutional object detector?

a. Softmax Layer

b. Linear Layer

c. Sigmoidal Layer

d. Absolute Value Layer

answer: b
reason: Classification의 output layers는 softmax output layer이며 regression heads의 output layers는 linear output layer이다.

40. Prior anchor boxes are usually sampled at random in image space before being used in the output layers of an object detector.

a. True

b. False

answer: b
reason: Anchor box는 IOU 계산 의해 가장 적절한 box가 사용된다.

41. While training an object detector, the cross entropy is calculated for the negative anchors only.

a. True

b. False

answer: a
reason: Cross entropy는 두 값을 확률적으로 비교하는 log 함수이다. 확률은 0 이상 1 이하이며 log를 취했을 때 항상 음수가 나온다.

42. When training an object detection model, the regression loss has the form:

$$L_{reg} = \frac{1}{N_p} \sum_i p_i L_2(b_i, b_i^*)$$

where the L2 norm is computed for every member in the minibatch. For a positive minibatch members, the value of P_i is:

answer: 1
reason: p_i는 negative일 때 0, positive일 때 1이다.

43. During non-maximum suppression, the output bounding box list is sorted based on the value of every member's:

a. Regression loss

b. IOU with ground truth

c. Softmax output score

d. Position in image space

answer: c
reason: Object detection output layer를 통해 bounding box가 그려지면 NMS 과정에서 가장 스코어가 높은 box만 남기고 나머지를 제거한다.

44. In context of self-driving cars, the output of object detectors can be used as a prior to perform which of the following tasks? (Check all that apply.)

a. 3D object detection

b. Traffic light state estimation

c. Drivable space estimation

d. Object tracking

answer: a, b, d
reason: Object detection 이후에 Lidar 센서를 추가 사용하여 3D object detection을 할 수 있으며 object 움직임에 대한 traking을 하고 교통신호 이해를 할 수 있다.

45. One of the main advantages of using the output of 2D object detectors as a prior to 3D object detection is their ability to easily handle occlusion and truncation.

a. True

b. False

answer: b
reason: 2D object detector만으로 occlusion과 truncation을 처리하기 어렵다.

46. Sudden camera motion is detrimental to the performance of object trackers. This is because tracking usually assumes gradual change in the camera's pose relative to the scene.

a. True

b. False

answer: a
reason: Object tracking assumption 세 가지로 '카메라는 갑자기 새로운 시점으로 이동하지 않는다', '물체는 사라지거나 다시 나타나지 않는다', '카메라가 움직인다면 점진적으로 움직인다'가 있다.

47. Achieving smooth category boundaries is a major difficulty to take into account while designing semantic segmentation models. Which of the following statements describe the origins of this problem? (Check all that apply.)

a. Thin objects such as poles, tree trunks, and lane separators.

b. Objects within the same category having variable appearances. An example being multiple color and models for cars on the road.

c. The similarity in appearance between some categories such as road, curb, and sidewalk.

answer:
reason:

48. When comparing the results of a semantic segmentation model to the ground truth, you found out that for the car category, its class IOU is 0.75. Knowing that the number of false positives (FP) is 17, and the number of false negatives (FN) is 3, what is the number of true positives achieved by this model?

answer:
reason:

49. To measure the performance of a semantic segmentation model over all classes, a good idea would be to average the class IOU.

a. True

b. False

answer:
reason:

50. Which of the following do you typically see in a Semantic Segmentation Model? (Check all that apply.)

a. Up-sampling layers in the decoder stage of the architecture.

b. Multiple Convolutional layers followed by an up-sampling layer.

c. Multiple Convolutional layers followed by a Pool layer.

d. Up-sampling layers in the encoder stage of the architecture.

answer:
reason:

51. Anchor boxes are an essential component of any semantic segmentation neural network architecture.

a. True

b. False

answer:
reason:

52. In your semantic segmentation model an input feature map is passed through a nearest neighbor up-sampling layer. The output feature map's depth is equal to that of the input feature map.

a. True

b. False

answer:

reason:

53. A standard semantic segmentation architecture that uses a softmax output layer is allowed to associate multiple categories to a single pixel in the input image.

a. True

b. False

answer:
reason:

54. Which of the below loss functions is usually used to train semantic segmentation models?

a. Mean Square Error (L2-Loss)

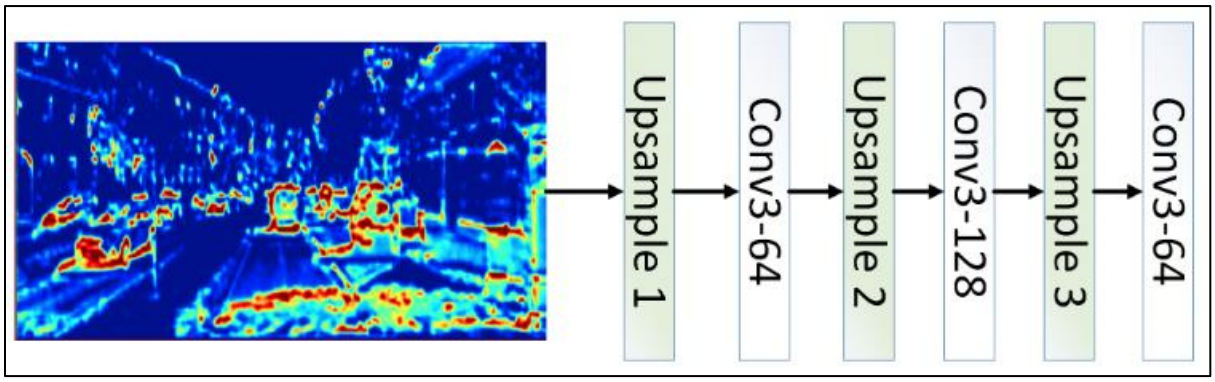
b. Cross-Entropy Loss

c. 0-1 Loss

d. Mean Absolute Error (L1-Loss)

answer:
reason:

55. A semantic segmentation model uses the following decoder architecture. The convolutions are all 3x3, have a padding size of 1, and have a number of filters shown in the figure. The up-sampling multiplier S is 2 for all upsampling layers.



If you pass an input of dimensions MxNxD through this decoder, what are the expected output dimensions?

Note: M is the width, N is the height, and D is the depth of the input.

a. 8*M, 8*N, 128*D

b. 6*M, 6*N, 64*D

c. 8*M, 8*N, 64*D

d. M/8, N/8, 64*D

answer: reason:

56. In context of self-driving cars, semantic segmentation can be used to perform: (Check all that apply.)

a. Drivable space estimation.

b. Velocity estimation of dynamic obstacles in the scene.

c. Constrain the image space used to perform 2D object detection.

d. Localization in a predefined 3D map.

e. Lane boundary estimation.

answer:
reason:

57. Which of the following categories in a semantic segmentation output map would be useful to determine lane boundaries? (Check all that apply.)

a. Road

b. Sidewalk

c. Curb

d. Lane Separator

e. Pedestrian

answer:
reason:

58. To estimate a plane model, an algorithm would require a minimum of:

a. Five points, chosen at random.

b. Three points, chosen to be non-collinear.

c. Three points, chosen to be collinear.

d. Five points, chosen to be non-collinear.

answer:
reason:

59. To estimate lines that could belong to lanes in a post-processed output image from semantic segmentation, containing only relevant categories, one would:

a. Use RANSAC to estimate the road plane, then fit lines to its boundary.

b. First apply Canny edge detection followed by a Kalman Filter to estimate lines.

c. First apply Hough transform line estimation followed by Canny edge detection.

d. First apply Canny edge detection followed by Hough transform line estimation.

answer:
reason: