1. **What do you mean by computer vision?**
   Computer Vision uses images and videos to understand a real-world scene. Just like Humans use eyes for capturing light, receptors in the brain for accessing it, and the visual cortex for processing it. Similarly, a computer understands images, videos, or a real-world scenario through machine learning algorithms and AI self-learning programming.

2. **What are some of the machine learning algorithms available in OpenCV?**
   Some machine learning libraries available in OpenCV are Artificial Neural Networks, Random Forest, Support Vector Machine, Decision Tree Learning, Convolution Neural Networks, Boosting and Gradient Boosting Trees, Expectation-Maximization algorithm, Naive Bayes classifier, K-nearest neighboring algorithm.

3. **How many types of image filters are in OpenCV?**
   Image filters used in OpenCV are:
   - i. Bilateral Filter
   - ii. Blur, Box Filter
   - iii. Dilate
   - iv. Build Pyramid
   - v. Erode
   - vi. Filter2D
   - vii. Gaussian Blur
   - viii. Deriv, and Gabor Kernels
   - ix. Laplacian
   - x. Median Blur.
4. **What is a face recognition algorithm? List some popular ones.**
   The face recognition algorithm is basically the computer application that is used for tracking, detecting, identifying, or verifying human faces simply from the image or the video that has been captured using the digital camera. Some popular but evolving algorithms are:
   - i. PCA- Principal Component Analysis
   - ii. LBPH- Local Binary Pattern Histograms
   - iii. k-NN (nearest neighbors) algorithm
   - iv. Eigen’s faces
   - v. Fisher faces
   - vi. SIFT- Scale Invariant Feature Transform
   - vii. SURF- Speed Up Robust Features

5. **What are some of the programming languages supported by Computer vision?**
    a. LISP,
    b. Prolog,
    c. C/C++,
    d. Java
    e. Python.
6. **What is a color model?**

    A Color Model is a coordinate system and a subset of visible colors. With "Color Model" we create a whole range of colors from a limited set of primary colors like RGB (Red Green Blue). Color Models are of two types: Additive and Subtractive.

7. **What Is Dynamic Range?**

    Dynamic Range is a ratio of small and large values that is assumed by a certain quantity. It is used in signals, photography, sounds, and light. From a photographic point of view, it is a ratio of minimum and maximum measuring light intensity or the lightest and darkest regions also called color contrast.

8. **What is Digital Image?**

    A digital image is an image that is comprised of the elements of the picture, they are also admitted as pixels. Each pixel is with the finite and the discrete numbers of the numerical representation which belong to its intensity and the gray level which is considered as its output from the functions of the two dimensions that are fed by the input of spatial coordinates that are denoted by the x-axis and y-axis.

9. **What are Erosion and Dilation?**

Erosion and Dilation are the two very common morphological image processing operations which is a procedure for modifying the geometric structure in the image.

10. **Which method is used to read images in OpenCV Python?**

The *cv2.imread()* method of the Imgcodecs class is used in OpenCV to read an image.

11. **In OpenCV which function is used to draw lines?**

    In OpenCV line() function is used to draw the line.

12. **Which method of OpenCV is used to save images?**

    cv2.imwrite() method is the method of OpenCV that is used to save an image.

13. **What is Haarcascade?**

    Haar cascade is an algorithm used for detecting or identifying the objects in the other images. It is a machine learning approach.

    In the Haar cascade there are two types of images:-

    1. Positive Image
    2. Negative Image

    The algorithm has four stages:-

    1. Haar Feature Selection
    2. Creating Integral Images
    3. Adaboost Training
    4. Cascading Classifiers

14. **Where you have hosted your Computer Vison model?**

    Computer vision applications can be deployed -:

    1. AWS
    2. Azure
    3. GCP

15. **What is Azure Computer Vision APIs?**

    The cloud-based Computer Vision API provides developers with access to advanced algorithms for processing images and returning information. By uploading an image or specifying an image URL, Microsoft Computer Vision algorithms can analyze visual content in different ways based on inputs and user choices.

16. **What was your frame per second?**

    Frames per second (fps), describes how smoothly a given game runs on your PC. The more frames you can pack into one second, the more smooth motion will be on-screen. Lower framerates that is, framerates lower than 30fps or so will appear choppy or slow.

17. **What was the data labeling tool that you have used for your project?**

    The data labeling tool that we have used are -:

    1. CVAT is an open-source data labeling tool and is also hosted in the cloud.
    2. Labelme
    3. LabelBox

18. **Explain me some of real-life use case of segmentation.**

    a. Medical Imaging

    b. Self-driving cars

    c. Satellite Imaging/ Remote sensing

19. **Explain types of Segmentation.**

   a. **Semantic segmentation**

      Semantic segmentation describes the process of associating each pixel with a class label. So simply, here we just care about a coarse representation of all the objects present in the image.

   b. **Instance segmentation**

      Unlike semantic segmentation, in image segmentation, we mask each instance of an object contained in an image independently. So this implies, that we will focus on the object of importance first and then identify each instance of the object separately.

   c. **Panoptic segmentation**

      When you combine semantic segmentation and instance segmentation, you get panoptic segmentation.

20. **Can you explain a scenario where you might use anchor boxes?**

    Anchor boxes are primarily useful for object detection. This means that a professional uses them to isolate different elements of an image, such as the size, shape and location so they can give those features values. I would use one in a situation where the image I want the computer to visualize includes a wide range variables.

21. **Explain what the mach band effect is?**

    The mach band effect is the optical illusion that happens when the edges of two images have similar shades of grey and the eye adjusts by interpreting a higher contrast between the two than there actually is. In computer vision, this phenomenon may cause inaccurate calculations. To account for these situations, I adjust the smoothness to reduce the banding effect a computer might detect.

22. **What are the drawbacks of VGGNet?**

    There are two major drawbacks with VGGNet:

    1. It is painfully slow to train.
    2. The network architecture weights themselves are quite large (concerning disk/bandwidth).

23. **Clarify with an example why the inputs in computer vision issues can get huge. Give a solution to overcome this challenge.**

Ans: Consider a 500x500 pixel RGB image fed to a fully connected neural network for which the first hidden layer has just 1000 hidden units. For this image, the number of input features will be 500*500*3=750,000, i.e. the input vector will be 750,000 dimensional. The weight matrix at the first hidden layer will therefore be a 1000x750,000 dimensional matrix which is huge in size for both computations as well as storage. We can use convolution operation, which is the basis of convolutional neural networks, in order to address this challenge.

24. **What are the features likely to be detected by the initial layers of a neural network utilized for Computer Vision? How is this different from what is detected by the later layers of the neural network?**

Ans: The earlier layers of the neural network detect simple features of an image, such as edges or corners. As we go deeper into the neural network, the features become increasingly complex, detecting shapes and patterns. The later layers of the neural network are capable of detecting complex patterns such as complete objects.

25. Consider a filter [-1 -1 -1; 0 0 0; 1 1 1] used for convolution.What edges will this filter extract from the input image?

Ans: This filter will extract horizontal edges from the image. To get a more concrete understanding, consider a grayscale image represented by an array with the following pixel intensities:

[0 0 0 0 0 0; 0 0 0 0 0 0; 0 0 0 0 0 0; 10 10 10 10 10 10; 10 10 10 10 10 10;].

From the array, it should be apparent that the top half of the image is black, whereas the lower half is a lighter color forming an apparent edge at the center of the image. The convolution of the two will result in the array [0 0 0 0; 30 30 30 30; 30 30 30 30; 0 0 0 0;]. It can be observed from the values in the resultant array that the horizontal edge has been identified.

26. How do you address the issue of the edge pixels being used less than the central pixels during convolutional operation?

Ans: In order to address the issue of the filter or kernel extracting information from the edge pixels less in comparison to the central pixel, we can use padding. Padding is essentially adding one or more additional rows or columns of pixels along the boundary of the image. The padding forms the new edge pixels of the image and therefore results in insufficient extraction of information from the original edge pixels. Padding provides the added advantage of preventing the shrinking of an image as a result of the convolution operations.

27. For a 10x10 image used with a 5x5 filter, what should the padding be in order to obtain a resultant image of the same size as the original image?

Ans: For an image without padding, the size of the image n x n after convolution with a filter of size f x f is given as (n-f+1)x(n-f+1) (You can verify this with the example in question 3. of this section).
Now, if you add padding of 1 pixel, in effect, this will be like adding a border of one pixel around an image, consequently increasing the length and breadth of the image by 2. That is, for a padding p, the size of the original image becomes (n+2p)x(n+2p). To maintain the size of the image after convolution, you will need to choose p=f-12.
For the given problem, therefore, we will need to choose p=(5-1)/2 =2

28. Given a 5x5 image with a 3x3 filter and a padding p=1, what will the size of the resultant image be if a convolutional stride of s= 2 is used?

Ans: For an nxn image with an fxf, padding p, and stride length s, the size of the resultant image after convolution has the shape n+2p-fs+1 x n+2p-fs+1. Therefore for the given problem the size of the resultant image will be (((5+2*1-3)/2) +1) x (((5+2*1-3)/2) +1)= 3 x 3.

29. For an RGB image of dimensions 10x10x3 convolved with a 3x3 filter, what will be the size of the resultant image?

Ans: The convolution operation is not possible for a 10x10x3 image with a 3x3 filter as the third dimension (or the number of channels) must be the same in order to achieve convolution. Alternatively, if a 10x10x3 image is convolved with a 3x3x3 filter, the dimensions of the resultant image will be 4x4.

30. What will be the change in the size of the results of the convolution is performed with a 1x1 filter on a 10x10 image with no padding?

Ans: The size of the image will remain unchanged; that is, the height and width of the resultant image will be the same as the resultant image. However, the number of channels in the resultant can be changed with a 1x1 convolution, thus increasing or decreasing the dimensionality of the input.

31. In order to extract multiple features from an image, it is common practice to use a number of filters and then stack them up to form the resultant. So for a 10x10x3 image convoluted using ten filters of dimensions 5x5x3, how many parameters must be learned to form the filters alone?

Ans: Each 5x5x3 filter has 5*5*3=75 features. Therefore for ten such filters, the total number of features will be 75*10=750 features. (NOTE: The bias has not been accounted for in this calculation. However, notice how this number is very small in comparison to the extremely large number of features mentioned in question 1 when convolution is not used.)

32. How many parameters are to be learned in the pooling layers?

Ans: No parameters are to be learned in the pooling layers. In general, the pooling layer has a set of hyperparameters describing the filter size and the stride length, which are set and work as a fixed computation.

33. For max-pooling done on a 6x6x3 image with a filter of size f=2 and padding p=0 with stride s=1, what would be the size of the output image?

Ans: The size of the resultant image will be as per the formula n+2p-fs+1. Additionally, the number of channels in the resultant image will be the same as that in the input image.

Therefore, for the given problem, the dimensions of the resultant image will be

(6+2*0-2+1) x (6+2*0-2+1) x 3 = 5 x 5 x 3.

34. Suggest a way to train a convolutional neural network when you have a quite small dataset.

Ans: If you have a dataset that will not be enough to train a convolutional neural network sufficiently well (and even otherwise sometimes in the interest of time or resource), it is suggested to use transfer learning to solve your machine learning problem. Depending on the size of the dataset and the budget of time and resources you have at your disposal, you can choose to train only the last classification layer or a few of the later layers. Alternatively, you could use transfer learning to train (when the size of the data set is large enough to allow this) all the layers with the parameters initialized from a previously trained model. Owing to the large availability of open-source pre-trained models, transfer learning for computer vision problems is both easy and strongly advised.

35. Explain why mirroring, random cropping, and shearing are some techniques that can help in a computer learning problem.

Ans: There is often limited data available to solve computer vision problems, which just isn't enough to train the neural networks. Techniques like mirroring, random cropping, and shearing can help augment the existing dataset and create more training data from the existing data, thereby ameliorating the issue of limited training data.

36. Mention a method that can be used to evaluate an object localization model. How does it work?

Ans: Intersection over Union (also known as IoU) is a commonly used method to evaluate the performance of the object localization model. The overlap between the ground truth bounding box and the predicted bounding box is checked, and the ratio of the intersection of the areas to the Union of the areas is calculated. If this ratio called IoU is found greater than some threshold (usually set to 0.5 or higher), the prediction of the model is considered correct.

37. How can IoU be used for resolving the issue of multiple detections of the same object?

Ans: The IoU method also helps in the non-max suppression technique for eliminating multiple detections of the same object. It is common for object localization models to predict multiple bounding boxes for the same object. In such cases, the box with the maximum probability among the overlapping boxes of a class is taken. The IoU of this box with the other overlapping boxes is then checked. If this IoU is above a certain threshold, the boxes are considered to be detecting the same object, and the lower probability boxes are eliminated.

38. Mention a scenario that would require the use of anchor boxes.

Ans: When detecting multiple classes of objects, there is a scenario where the center of two bounding boxes capturing objects of two different classes occurs on the same point or grid cell. Despite the overlap of the two boxes owing to the fact that both the boxes indicate different objects, they will both require to be retained (something a grid cell in the sliding window technique isn't ordinarily able to accomplish). For this purpose, a number of anchor boxes of different dimensions are used, and vectors similar to the original output vector of a grid cell are now provided for each anchor box in the new output vector. The best-fitting anchor box for a particular class of objects is used to indicate that the grid cell contains the center of the object, thus allowing for multiple overlapping object detection.

39. How does the Siamese Network help to address the one-shot learning problem?

Ans: Siamese Network works by encoding a given image with a learned set of parameters such that the distance between the encoding is large when the image is of two different people, and the distance is small when the two images compared are of the same purpose. Training this neural network, therefore, involves learning parameters such that these conditions of the encoding are satisfied when provided with images of two different people or two images of the same person.

40. What purpose does grayscaling serve?

Ans: Grayscaling helps to reduce the dimension of the image and thus allows for reduced computation time and effort. Further, it reduces the complexity of models and functions required for various operations. Some functions like edge and contour detection and machine learning problems Optical Character Recognition perform better or are implemented for working only with grayscale images.

41. What color to grayscale conversion algorithm does OpenCV employ? What is the logic behind this?

Ans: The color to the grayscale algorithm in OpenCV uses the formula Y=0.299*R+0.587*G+0.114*B. This makes it similar to the luminosity method, which averages the color intensity values weighting them in accordance to human perception of different colors, i.e., it accounts for the fact that humans perceive green more strongly than red, and red more strongly than blue, which is apparent from the weightage given to each color's pixel intensity. Additionally, the OpenCV grayscaling algorithm takes into consideration the nonlinear operation used to encode images.

42. What is translational equivariance? What brings about this property in Convolutional Neural Networks?

Ans: Translational Equivariance is a property where the position of an object in an image will not affect its detection in the image. That is, if an image is shifted a few pixels to the right or to the left, the output will merely change its position equally and otherwise remain unaffected. Translational Equivariance is an important property of Convolutional neural networks and is brought about by the parameter sharing concept.

43. What is the basis of the popular EAST text detector?

Ans: EAST text detector is based on a fully convolutional neural network adapted for text detection. It has gained immense popularity because of its text detection accuracy in natural scene images.

44. What is the basis of the state-of-the-art object detection algorithm YOLO?

Ans: YOLO is an object detection algorithm that is based on a Convolutional Neural Network and capable of working in real-time. It provides accurate detections for a large variety of objects and can be used as a solution or a starting point for transfer learning for many computer vision problems.

45. Can you name some of the different types of image filters used in OpenCV?

Ans: Gaussian Blur, Bilateral Filter, Gabor Kernels, Median Blur, Dilate, Box Filter, and Erode.

46. Can you define digital image?

Ans: A digital image is a picture that's made up of smaller parts, called pixels. These pixels are made of numerical components that represent their color codes and intensity. AI systems use these numbers to understand an image.

47. What programming languages does computer vision support?

Ans: Computer vision can use programming languages such as Java, C/C++, Prolog, Python, and LISP. I've primarily used C++ in past projects, but I have certification in Python and a basic understanding of the others.

48. Match the following image formats to their correct number of channels

- GrayScale
- RGB
  - I. 1 channel
  - II. 2 channels
  - III. 3 channels
  - IV. 4 channels

None
A) RGB -> I, GrayScale-> III
B) RGB -> IV, GrayScale-> II
C) RGB -> III, GrayScale -> I
D) RGB -> II, GrayScale -> I

Solution: C:
Grayscale images have one number per pixel and are stored as an m × n matrix, whereas Color images have 3 numbers per pixel – red, green, and blue brightness (RGB)

49. [True or False] To blur an image, you can use a linear filter

A) TRUE
B) FALSE

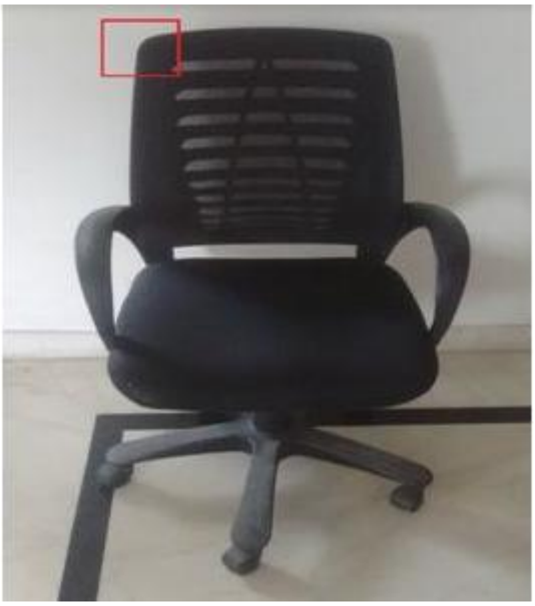
Solution: B
Blurring compares neighboring pixels in a filter and smooths them. For this, you cannot use a linear filter.

50. Which of the following is a challenge when dealing with computer vision problems?

A) Variations due to geometric changes (like pose, scale etc)
B) Variations due to photometric factors (like illumination, appearance etc)
C) Image occlusion
D) All of the above

Solution: D
All the above-mentioned options are challenges in computer vision

51.



In this image, you can find an edge labeled in the red region. Which form of discontinuity creates this kind of edge?

A) Depth Discontinuity
B) Surface color Discontinuity
C) Illumination discontinuity
D) None of the above

Solution: A
The chair and wall are far from each other, causing an edge in the image.

52. Finite difference filters in image processing are very susceptible to noise. To cope up with this, which of the following methods can you use so that there would be minimal distortions by noise?

A) Downsample the image
B) Convert the image to grayscale from RGB
C) Smooth the image
D) None of the above

Solution: C
Smoothing helps in reducing noise by forcing pixels to be more like their neighbors

53. [True or False] Quantizing an image will reduce the amount of memory required for storage.

A) TRUE
B) FALSE

Solution: A
The statement given is true.

54. Suppose we have a grayscale image, with most of the values of pixels being the same. What can we use to compress the size of the image?

Ans: Encoding the same values of pixels will greatly reduce the size of the storage

55. [True or False] JPEG is a lossy image compression technique?

A) TRUE
B) FALSE

Solution: A

The reason for JPEG being a lossy compression technique is because of the use of quantization.

56. Which of the following methods is used as a model-fitting method for edge detection?

A) SIFT
B) Difference of Gaussian detector
C) RANSAC
D) None of the above

Solution: C
RANSAC is used to find the best fit line in edge detection

58. Suppose you are creating a face detector in the wild. Which of the following features would you select for creating a robust facial detector?

1. Location of iris, eyebrow, and chin
2. Boolean feature: Is the person smiling or not
3. Angle of the orientation of the face
4. Is the person sitting or standing

A) 1, 2
B) 1, 3
C) 1, 2, 3
D) 1, 2, 3, 4

Solution: B
Options 1, 3 would be relevant features for the problem, but 2, 4 may not be

59. Which of the following is an example of a low-level feature in an image?

A) HOG
B) SIFT
C) HAAR features
D) All of the above

Solution: D
All the above are examples of low-level features

60. In the RGBA mode of color representation, what does A represent?

A) Depth of an image
B) Intensity of colors
C) Opacity of an image
D) None of the above

Solution: C
Opacity can be mentioned by introducing it as the fourth parameter in RGB

61. Which of the following data augmentation technique would you prefer for an object recognition problem?

A) Horizontal flipping
B) Rescaling
C) Zooming in the image
D) All of the above

Solution: D
All the mentioned techniques can be used for data augmentation.

62. How many types of image filters in OpenCV ?

Ans:

Averaging
Gaussian Filtering
Median Filtering
Bilateral Filtering

63. Popular computer vision libraries

Ans:

OpenCV.

- SimpleCV.
- TensorFlow.
- Keras.
- MATLAB.
- PCL.
- DeepFace.
- NVIDIA CUDA-X.
- Pytorch

64. What does it mean to use "inception architecture" for a CNN? What does it solve?

An inception network is a deep neural network with an architecture that consists of inception modules, which are repeating components. Using inception architecture introduces inception blocks, which contain multiple convolutional and pooling layers stacked together. This gives more accurate results and can help to reduce computation costs.

65. How would you encode a categorical variable with thousands of distinct values?

The approach to this encoding categorical features question depends on whether the problem is a regression or a classification model. If it's a regression model, one possible solution would be to cluster them based on the response by working backwards. You could sort them by the response variable and then split the categorical variables into buckets based on the grouping of the response variable.