

IITM – Arihant CV Apps Assignment

Q1. Identifying the number of glass sheets in the given image.

Following is the sequence of steps that I took to solve the above problem:

1. Given we are counting the number of glass sheets and glass being a transparent/translucent or reflective material, color is not a property that can be associated with the glass sheets. Hence, to reduce the redundant info, convert the given image from RGB map to Grayscale color map.
2. Do a standard noise filtering to avoid spurious results. However, given the problem statement of counting the number of sheets, vertical / horizontal edge detection algorithm can help with the task. So, we need to preserve the high frequency info. Hence, instead of blurring with averaging filter, filter the image using median filter to remove any salt and pepper noise.
3. Given the reflections of ceiling lights are creating a saturated region within the image of the glass sheets, apply homomorphic filter to filter out the illumination component, and retain only the reflectance component of the image.
4. Since the glass sheets are stacked up horizontally, vertical edge detection can help in finding the boundary between 2 sheets. Hence, apply Sobel_X kernel filter to extract vertical edges.
5. Finally, to localize the edges better, non-maximal suppression needs to be done as implemented in Canny edge detector. However, approximate results can also be obtained using the in-built erode filter of OpenCV toolbox. Hence, apply the erode filter to the derivative image. This resultant image can be used to count the number of glass sheets either by counting manually or through Hough line detection algorithms.

Q2. Deep Learning Model to classify images

Architecture used: VGG_16. (FC layers modified to meet the requirements of the given problem).

Following are the reasoning/intuition behind the model modification:

1. No. of output classes has changed from ImageNet 1000 classes to 6 classes.
2. Given all the image categories belong to air vehicles (drones, planes, etc.), the pen-ultimate FC layer of 4096 -> 4096 mapping doesn't make much sense, rather the ability to learn how various features are connected to each other will help better classify the images.

Apart from model modification, for training, the images have been augmented using random crop and flip to prevent overfitting and enable generalization.

For training, 96 batches of size 32 images for each class has been chosen, and a validation set of 64 images for each class has been chosen.