CS 783A: Visual Recognition (Mid-Semester Examination)

Vinay P. Namboodiri

19 February 2019 (8 am - 10 am)

Total Number of Pages: 4

Total Points 30

Question	Points	Score
1	5	1.5
2	5	0.5
3	5	2
4	5	3+2
5	5	3
6	5	3
Total:	30	13+2

15

Instructions

1. Read these instructions carefully.
2. Write you name, section and roll number on all the pages of the answer book, **including the ROUGH pages**. You will be penalised if you fail to write the name, roll number and correct section.
3. Write the answers cleanly in the space provided. Space is given for rough work in the answer book.
4. Using pens (blue/black ink) and not pencils. Do not use red pens for answering.
5. Do not exchange question books or change the seat after obtaining question paper.
6. Even if no answers are written, the answer book has to be returned back with name and roll number written.
7. Sign the attendance sheet.
8. No clarifications will be provided. Make suitable assumptions and specify your assumption in the paper.

Question 1. (5 points) For instance recognition using video google with a database of 1000 images, we initially obtained a bag of words using 100 cluster centers. Using these, we created an initial inverted file list. However, the database was later changed and 500 more images were added. Using the 500 additional images, we obtain 100 more cluster centers. How would you modify the algorithm to make use of the additional 100 cluster centers?

Answer:

Assuming the new images, and no change in cluster assignment of old image. new cluster and mapped to sift features in word ≃ cluster centre.

Basically, there are some new words coming. New images contains new word and old word. So in bag of word representation, we increase the no of bags to (100 + 100) = 200. For old image, no of new word will be zero. For each new image we create a new BoW histogram (~~only contain~~ both new and old word number may be non zero).

In inverted file index, we append the new image to the old word index, if new image contain that old word. Do the new file inverted index for new words.

1.5

Question 2. (5 points) While obtaining Harris corner detector, we ensured that the points obtained were distinctive. Using this we obtained a set of K keypoints for an image. If we would like to obtain 10 more keypoints what change would we make?

Answer:

In harris corner we use $\frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} = \frac{\det(H)}{\text{trace}(H)}$ and select a keypoint if it is bigger than a threshold. we would decrease that threshold.

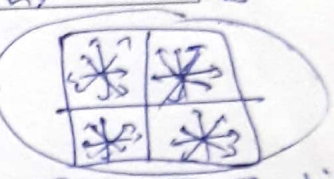
0.5

Question 3. (5 points) In SIFT descriptor, how do we assign orientation to keypoints. Are all keypoints assigned unique orientations? If yes, why, if not, why not?

Answer:

We calculate gradient magnitude

$m = \sqrt{\left(\frac{dG}{dx}\right)^2 + \left(\frac{dG}{dy}\right)^2}$ and its direction slope $= \frac{dG}{dy} / \frac{dG}{dx}$. we consider 8 direction and assign its direction to one of 8 (to the nearest). All points in neighbour of keypoint are assigned orientation this way. No, orientation is not unique, it is one of 8.



sift descriptor consist of 4 such orientation. Each orientation is a sum of orientation of various pixel.

2

Question 4. (5 points) In the pyramid match kernel we ensure that the matching is done once and not repeated at all levels. How do we achieve this? At what level is the matching counted?

Answer:

kernel $= \sum w_i N_i$

So, N_i = Intersect(level i) − Intersect(level $i-1$)

So, we are substracting the intersections done at lower level and thus it is counted ones.

For ith level Histogram, residual matchings is counted in N_i where $K = \sum_i w_i N_i$.

3 + 2

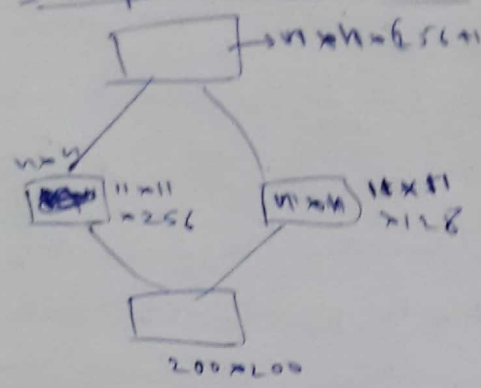
Name:

Rollno:

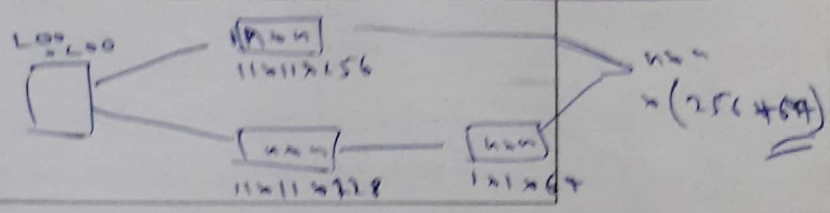
Question 5. (5 points) In convolutional neural networks, one way to ensure reduced set of parameters is by using 1x1 convolutions. For instance, this is used in inception network. How does a 1x1 convolution result in reduced set of parameters?

Answer:

Inception network



1×1 works like fully connected layer and it reduces no of parameter and reducing the number of filter / channels.



3

Question 6. (5 points) In the Histogram of Oriented Gradients (HoG) feature, the authors propose use of both cells and blocks. What is the difference between cells and blocks. What is obtained in each cell and block? ?

Answer:

Each image is divided into cells ~~oriented~~ directions like (6×6) and it consists of histogram of gradients. while each block consists of cells like (3×3) and cossists of various cells and their histogram. Also Two block can overlap on each other. Cell represent a basic entity feature in image, while block recognise bigger features like legs (using aggregate of basic features).

3