CS376: Computer Vision: Assignment 5
Due: May 19th, 11:59 PM

1 Short answer problems [20 pts]

1. What is the relation between image classification and object detection? Please give two concrete examples on how they are related (e.g., one is applied to solve another)

2. Please compare semantic segmentation, instance segmentation and object detection. Describe the similarities and differences.

3. Describe at least two plausible neural network designs for the task of semantic segmentations.

4. What are the applications of generative models in solving core computer vision tasks.

2 Face Detection (40 points)

The first programming assignment is to implement a face detector in natural images. We will follow the standard pipeline of sliding window + classification. For classification, please implement either a boosting framework or a SVM framework. For the boosting framework, please follow the Viola-Jones algorithm:

- Rapid object detection using a boosted cascade of simple features. CVPR 2001. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.6807. Note that there are many other resource to read, such as https://en.wikipedia.org/wiki/Viola--Jones_object_detection_framework.

For the SVM framework, please use the HOG feature representation used in assignment 4.
For both programming assignments, please stick to the programming environment that you used in assignment 4.
We will use the Face Detection Data Set and Benchmark Home http://vis-www.cs.umass.edu/fddb/. Regarding training/testing, we will use the first 9 folders for training and the last 2 folders for testing. Details can be find in the readme file of FDDBhttp://vis-www.cs.umass.edu/fddb/README.txt:

- Training: FDDB-fold-00, FDDB-fold-01, $\cdots$, FDDB-fold-08;
- Testing: FDDB-fold-09, FDDB-fold-10;

Note that this dataset annotates a face using an ellipse. Please convert an ellipse into a rectangular box. One approach can be found athttps://stackoverflow.com/questions/87734/how-do-you-calculate-the-axis-aligned-bo
In terms of evaluation, please report the precision-recall curve with an IOU threshold of 0.5.
Please submit the following:

- **Face classification (20 points):** Please implement an approach that takes an image patch as input, and output whether it contains a face in it or not. Note that how to generate the training data is important. For negative instances, please include non-face patches and faces that overlap with human objects but are too big or too small. We will look into how you train the classifier and how you generate the training data. Please report the classification accuracy of your approach.
- **Sliding window + Human classification (20 points):** Implement a sliding window approach and combine human classification to perform face object detection. Please report the precision-recall curve with an IOU threshold of 0.5 on 'sec06'.

3 Human Detection (40 points)

We would like to extend face detection to human subjects (pedestrians). You will use the same boosting/SVM framework. We will use the Caltech pedestrian detection dataset http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/datasets/USA/. For simplicity, we will use 'set00' for training and test on 'set06'. However, it is recommended to use all training data for the extra credit session, i.e., when you develop neural networks to detect pedestrians.

- **Human classification (10 points):** Please implement an approach that takes an image patch as input, and output whether it contains a human subject in it. Note that how to generate the training data is important. For negative instances, please include non-human patches and patches that overlap with human objects but are too big or too small. We will look into how you train the classifier and how you generate the training data. Please report the classification accuracy of your approach.
- **Sliding window + Human classification (10 points):** Implement a sliding window approach and combine human classification to perform human object detection. Please report the precision-recall curve with an IOU threshold of 0.5.
- **Efficient implementation (20 points).** Implement a speed-up approach for human object detection. If you use Boosting, you can use fewer weak-classifiers to perform early stop. If you use SVM, then you can train multiple SVMs with fewer features and use them to perform early stop. Please report the precision-recall curve with an IOU threshold of 0.5 and the running time before and after speeding up the calculation.

4 Extra Credits (50 points)

We are going to offer 50 extra credits for excellent algorithms and implementations that boost the performance of human detections in natural images. You are not restricted a particular approach. You can also use additional training data to train your object detector. However, the evaluation is always on testing set described above.

We are looking for creative solutions. Below are some ideas that you can implement:

- Train a neural network to classify each sliding window.
- Train a dense prediction network to regress the corners of the bounding boxes of the detected objects.
- Think about how to utilize the context, which is very important for detecting partially occluded human objects.

Please keep in mind that it is recommended to have unique ideas.

We will grade this programming assignment from the following four perspectives:

- **Object detection (25 points):** We will run a competition here. Please report the precision-recall curve with an IOU threshold of 0.5. For all submissions that improve from the non-deep learning approach above, we will compare the precision-recall curves. If your submission is ranked ith out of n total number of submissions, then you will get $20 + 5 * (1 - \frac{i-1}{n-1})$ points.
- **Running time (15 points):** We again run a competition here. For all submissions that improve from the non-deep learning approach above, we rank each submission based on the running time. If your submission is ranked ith out of n total number of submissions, then you will get $10 + 5 * (1 - \frac{i-1}{n-1})$ points.
- **Creative ideas (10 points):** We encourage creative ideas for performing object detection. You will get 5 points with an approach that is not discussed in class. You will get 10 points if your approach differs from what has been proposed in the literature.

Below are some useful papers to look at for human object detection using neural networks:

- Is faster r-cnn doing well for pedestrian detection? `https://arxiv.org/pdf/1607.07032.pdf`
- A unified multi-scale deep convolutional neural network for fast object detection. `https://arxiv.org/pdf/1607.07155.pdf`
- Occluded pedestrian detection through guided attention in CNNs. `http://openaccess.thecvf.com/content_cvpr_2018/papers/Zhang_Occluded_Pedestrian_Detection_CVPR_2018_paper.pdf`
- Cornernet: Detecting objects as paired keypoints. `http://openaccess.thecvf.com/content_ECCV_2018/papers/Hei_Law_CornerNet_Detecting_Objects_ECCV_2018_paper.pdf`

Submission instructions:

Create a single zip file so submit on Canvas that includes

- Your well-commented code, including the files and functions named as specified above.
- A pdf writeup of your results with embedded figures where relevant.

Please do not include any saved matrices or images etc. within your zip file.