Spring 2023 ELEC/COMP 447/546 Final Exam

Name:

NetID:

There are 7 problems in this exam (8 for ELEC/COMP 546), and several of them have multiple subparts. Remember to explain answers if the question prompt says “Explain.”

Grade (for instructor use only):

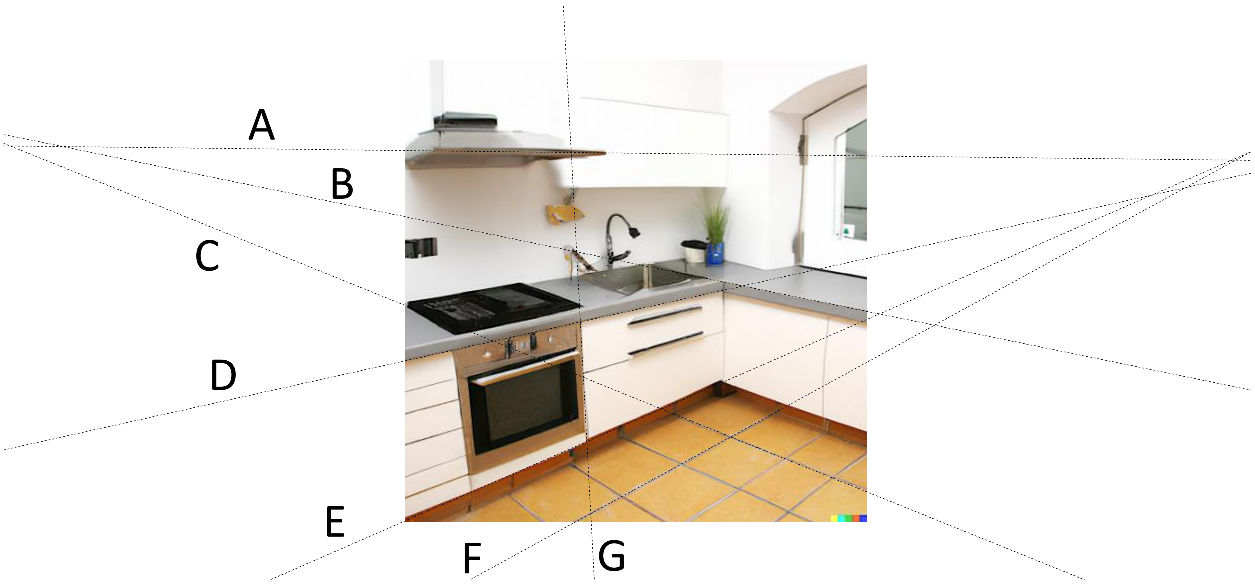
Problem	Earned Points	Total Points
1		15
2		15
3		5
4		5
5		40
6		10
7		10
8 (grad only)		20
Total		100 (grad: 120)

Problem 1: Cameras and Geometry

1.1 Circle all true statements about pinhole cameras:

a. The image is virtual.
b. The image is inverted in orientation.
c. The image is typically smaller in size than the actual scene.
d. The image is in grayscale.
e. The image is formed by refraction through a lens.
f. The image is formed by light from each world point traveling in a line.
g. The image gets sharper as the aperture becomes larger.
h. The image gets fainter as the aperture becomes smaller.
i. The smaller the screen, the smaller the field of view.

1.2 Your neighborhood realtor says there is a great house nearby that he wants to sell you and sends you a picture of the kitchen. You are suspicious the picture isn't real, and in fact generated by Dalle-2 (maybe because of the color strip in the bottom right corner). You decide to use your knowledge of camera geometry to determine if the picture is fake. You first draw various lines (A-G) overlaid on straight edges in the scene. How do you reason from these lines that the image is not real? Explain.



1.3 You are given two cameras with all the internal and external calibration information for both.

a. Suppose the two cameras differ by a translation factor. Can you transform an image from the point of view of one camera to an image from the point of view of the other camera? Explain.
b. Suppose the two cameras have the same center-of-projection and differ by a rotation factor. Can you transform an image from the point of view of one camera to an image from the point of view of another camera? Explain.

Problem 2: Signal Processing

2.1 An image is being filtered by a 5 x 5 kernel. You do not know the elements of the kernel. However, you know that each element was drawn independently from a standard Normal distribution. With only this information, is it more likely that this filter is a smoothing filter or an edge detection filter? Explain.

2.2 Which of these filters can be implemented as a convolutional operation over an image?

Median filter: $I(p) = \underset{q \in N(p)}{\operatorname{med}} I(q)$ (replaces pixel with the median value in a neighborhood)
Minimum filter: $I(p) = \min_{q \in N(p)} I(q)$ (replaces pixel with the minimum value in a neighborhood)
Bilateral filter: $I(p) = \sum_{q \in N(p)} G(p - q, \sigma_d) \cdot G(I(p) - I(q), \sigma_v) \cdot I(q)$

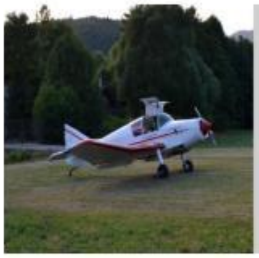
Where:

- $p = (p_x, p_y)$ and $q = (q_x, q_y)$ are pixel coordinates.
- $N(p)$ refers to the local spatial neighborhood around pixel coordinate p.
- $G(d, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-d^T d}{2\sigma^2}\right)$ is the Gaussian kernel.

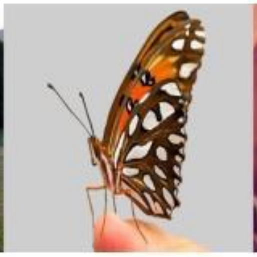
a. None of the above
b. Median filter
c. Bilateral filter
d. Minimum filter
e. Median and bilateral filter
f. All the above

2.3 Match the images below with their corresponding Fourier spectra magnitudes:

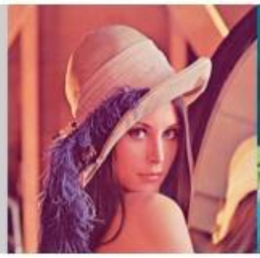
a



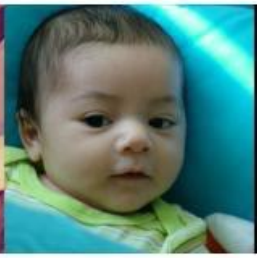
b



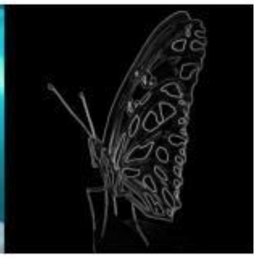
c

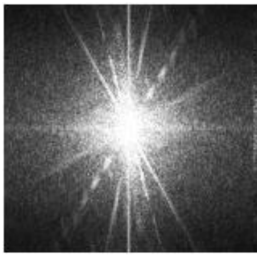


d

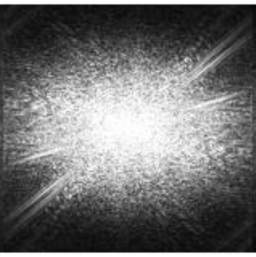


e

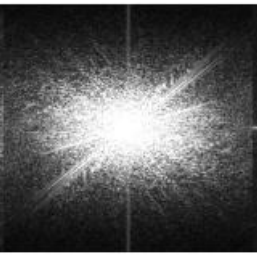




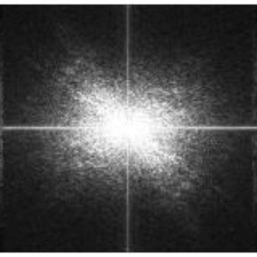
1



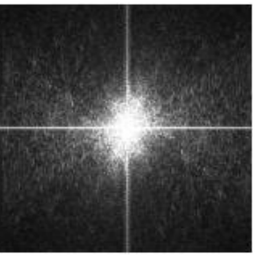
2



3



4



5

a: b: c: d: e:

Problem 3: Image Pyramids

Explain the type of tasks Gaussian pyramids are more useful for than Laplacian pyramids, and vice versa.

Problem 4: Optical Flow

Tiger Woods is trying to track feature points of a video of himself swinging a golf club using the Lucas-Kanade algorithm. He finds that while points track well on the club head (the part that hits the ball), they do not track well on the shaft. Explain why this happens mathematically and state the common name for this problem used by the vision community.

Problem 5: Machine Learning and Deep Learning

5.1 For which of these tasks would a bag-of-visual-words classifier be worst? Explain.

a. Classifying whether or not two people are looking at each other
b. Classifying the weather from an outdoor image
c. Classifying oak, maple, and cedar trees
d. Classifying texture patterns

5.2 Consider this simple 2-layer CNN that takes an RGB image as input:

Layer	Operation	Output channels
1	3 x 3 Conv + ReLU	10
	MaxPool (2 x 2)	10
2	3 x 3 Conv + ReLU	20

a. How many parameters does this model have?

b. How many filters are in the 2nd layer Conv?

c. What is the receptive field of the 2nd layer Conv's filters?

5.3 What advantage do residual ('skip') connections give in neural network training?

5.4 What is the difference between semantic and instance segmentation? Is a U-Net appropriate for either task? What about Mask-RCNN?

5.5 You are designing a pretext task to learn self-supervised features. Which of the following is the **least** likely to help learn useful features? Explain.

a. Predicting rotation angle for a rotated image
b. Inpainting masked regions of an image
c. Predicting grayscale values from an RGB image
d. Autoencoding an image

5.6 Which one of the following is **not** an image-to-image translation task?

a. Grayscale to RGB
b. RGB to Grayscale
c. Super-resolution
d. Denoising
e. Segmentation map to RGB

5.7 What bias does a CNN impose on its learned features that a vision transformer does not, and how does that help transformers outperform CNNs for many tasks?

5.8 Explain why GANs are often unstable to train.

Problem 6: Spot the coding errors

Your friend Dina is creating a face pose estimation model. The model will take an image of a face and output a rotation angle $\theta \in [-\pi, \pi]$ about the vertical axis (known as ‘yaw’). She has collected training data (RGB images with associated ground truth angles) and implemented the code. Hearing that you are a computer vision expert, she asks you to check her code. Identify three major errors in her model training code snippet below (the errors are **not** syntax-related):

```
batch_size = 64
transform = [T.ToTensor()]
transform.append(T.RandomHorizontalFlip())
transform.append(T.ColorJitter(brightness=(0.9,1.1)))
transform = T.Compose(transform)
train_dataset = PoseDataset('/data/PoseDataset/train', transform)
train_dataloader = data.DataLoader(train_dataset, batch_size=batch_size)

model = Model()
loss_fn = nn.CrossEntropyLoss()
optimizer = torch.optim.Adam(model.parameters(), lr=1e-4)

for epoch in range(10):
  model.train()
  for batch, (X,y) in enumerate(train_dataloader):
    pred = model(X)
    loss = loss_fn(pred, y)
    optimizer.zero_grad()
    optimizer.step()
```

Problem 7: System design
Your company is making a face recognition algorithm that will be used as a component of an automatic lock app. The app will allow or deny people entry into their homes at the front door based on whether the person’s face is in an accepted set of faces for that home (this set can be updated at any time by the homeowners). Comment on the different design aspects you must consider, including (but not limited to):

1. Dataset collection
2. Fairness/bias
3. Model design and computational cost
4. Evaluation (what metric(s) are important for this application?)

State any assumptions that you are making while answering the question.

Problem 8 (ELEC/COMP 546 ONLY)

8.1 Explain what the bilateral filter does based on the provided definition in Problem 2.2.

8.2 CNNs typically contain millions of parameters but are often able to be trained on datasets on the order of 10,000 images. Why then do CNNs not immediately overfit the data?

8.3 Why do we typically initialize weights of a neural network to small random values (as opposed to all zeros or large random values)?

8.4 CNN classifiers are prone to adversarial attacks, i.e., injection of imperceptible high-frequency details to images that change the classifier's output. If CNNs are affected by such small changes, why are they not as affected by the addition of slight Gaussian noise to images?