Sample Questions

YSC3221: Computer Vision and Deep Learning

Note: These questions below indicate only the "style" of the actual exam questions. They might be more or less difficult than the actual ones. The expectation is that you thoroughly understand the theoretical aspect of every topic in the course. For this, read the lecture notes and the reading material (if necessary the additional resources as well). Note that, the below questions focus on few selected topics. The final exam questions are taken across all topics we've learned.

1. Why are image features important in many problems in computer vision? Write 3 reasons why we need image features and not all pixels of the input image.

2. Explain how the integral image can solve the problem of arbitrary face size and unknown face locations.

3. Why does SIFT use Taylor expansion in one of its steps?

4. What is the main advantage of HoG over Haar-like features or SIFT?

5. Consider the following equation for an undirected graph:

$$\{x\}^* = \arg\max_{\{x\}} \prod_i^n p(\{x_i\}, \{d_i\}), \tag{1}$$

   where x_i is the latent variable at location i, and d_i is the observation.

   State briefly the meanings of:

   (a) $\{x\}^*$
   (b) $\arg\max_{\{x\}}$
   (c) $\prod_i^n p(\{x_i\}, \{d_i\})$

6. Consider Fig. 6.a and write the data and prior terms of an MRF, so that when you optimize the MRF based on the image, you can have an image similar to Fig. 6.b. Justify why your data and prior terms can clean the noise. You may assume the noise is modeled by the Gaussian noise.

7. Consider the following equations:

$$\begin{aligned} x_t^* &= \arg\max_{\{x_t\}} p(x_t|\{d\}_t) && (2)\\ &= \arg\max_{\{x_t\}} p(d_t|x_t) \sum_{\{x_{t-1}\}} p(x_t|x_{t-1}) p(x_{t-1}|\{d\}_{t-1}). && (3) \end{aligned}$$

(a) Draw the graphical model of Eq. 3.

(b) Show the detailed derivation from Eq. 2 to Eq. 3.

8. Regarding geometric camera calibration:

   (a) Discuss the basic reasons why we need a checkerboard to calibrate a camera.

   (b) Provide an example of other means (rather than a checkerboard) to calibrate a camera, and justify why it can work.

9. Below is the intrinsic camera matrix:

$$\mathbb{K} = \begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

   (a) What are x_0 and y_0? Justify your answer.

   (b) In some cases f_x and f_y can have different values. Why? And, what are the consequences?

10. The center of a camera can be computed from the extrinsic parameters.

   (a) Show the formula to compute the camera's center.

   (b) Explain why that formula works (in other words, show the geometrical meaning of the formula; if necessary draw a diagram to support your arguments).

11. To estimate a camera matrix, $\mathbb{P}$, we can pose the problem to minimizing the following cost function:

$$\begin{aligned} \min \quad & |\mathbb{A}p| \\ \text{s.t.} \quad & |p| = 1 \end{aligned} \tag{5}$$

   (a) Why do we need the constraint, i.e., $|p| = 1$.

   (b) We can solve the last objective function using SVD. However, we also need to refine the estimation further. Why?

12. Given camera matrix $\mathbb{P}$, write the step-by-step operations of backprojecting a pixel located at (x_i, y_i) to the 3D world.

13. Consider an MRF, where the latent variables denoted as x's and the observation denoted as d's.

   (a) Draw the graphical model for a 2x2 MRF that includes the latent variables and the observation.

   (b) Regarding the equation below, explain as detailed as possible about Z and ϕ. Why don't we use $p(x_i, d_i)$ and $p(x_i, x_j)$?

$$p(\{x\}), \{d\}) = \frac{1}{Z} \prod_{i}^{n} \phi[x_i, d_i] \prod_{j \in N_i} \phi[x_i, x_j]. \tag{6}$$

   (c) Discuss the meanings of factorization and conditional independence in relation to Eq.(6).

   (d) To estimate the values of the latent variables, x's, using MAP (Maximum A Posterior), we can use Eq.(7) below. Show the step-by-step derivation from Eq. 6 to Eq.(7).

$$\{x\}^* = \arg\min_{\{x\}} \left(\sum_{i}^{n} f_d(x_i, d_i) + \sum_{j \in N_i} f_p(x_i, x_j) \right) \tag{7}$$

14. In a neural network, if we use sigmoid functions for the activation functions of the neurons, then we apply the following equations for the backpropagation:

$$\delta_j^L = (a_j^L - y_j)\frac{e^-z_j^L}{(1+e^-z_j^L)^2}, \tag{8}$$

$$\delta_j^l = w_k^{l+1}\delta_k^{l+1}\frac{e^-z_j^L}{(1+e^-z_j^L)^2}, \tag{9}$$

$$\frac{\partial E}{\partial w_{jk}^l} = a_k^{l-1}\delta_j^l, \tag{10}$$

$$\frac{\partial E}{\partial b_j^l} = \delta_j^l. \tag{11}$$

(a) Write the stochastic mini-batch gradient descent equations for updating the weights and biases.

(b) Modify the above equations if we want to use ReLU functions, $f(x) = \max\{0, x\}$.

15. Show mathematically that the cross entropy error function (Eq.(12)) is generally faster than the quadratic error function in optimizing a neural network.

$$E(w,b) = -\frac{1}{n}\sum_x^n (y\log a + (1-y)\log(1-a)) \tag{12}$$

16. Discuss in what conditions deep learning can suffer from the overfitting problem. Also, explain in detail how and why adding a regularization term can reduce the problem.

17. If we replace the usual non-linear activation function $f(z) = \sigma(z)$ with $f(z) = z$ throughout the network. Rewrite the definitions of BP 1 and BP 2. Note that:

$$\delta_j^L = \frac{\partial E}{\partial z_j^L} \tag{13}$$

$$\delta_j^l = \frac{\partial E}{\partial z_j^l} = \sum_{m=0}^{M}\frac{\partial E_m}{\partial z_m^{l+1}}\frac{\partial z_m^{l+1}}{\partial z_j^l} \tag{14}$$

where $E(\mathbf{w},\mathbf{b}) = \frac{1}{2}\sum_n ||t_n - a_n||^2$.