Student Name: Mingqian Liao
Student ID: 3038745426

Computer Vision(CS280)
Homework1

---

1. Perspective Projection

   (1) We can define the point P in a plane of 3D space with normal vector n of the plane, then we could represent any point P_0 of the plane as $n(P_0 - P) = 0$ where λ is a scalar. Then consider a line L on the plane, and point P_1 is on the line, so we could represent the point as $P = P_1 + \lambda d$ where d is a unit direction vector, and λ is a scalar. Then we can substitude the line equation into the plane equation, then we can obtain

   $$n(P_0 - (P_1 + \lambda d)) = 0$$

   so that when $n * d \neq 0$,

   $$\lambda = \frac{n(P_0 - P_1)}{\lambda d}$$

   Substitute λ into the line equation, then we can obtain

   $$P = P_1 + \frac{n(P_0 - P_1)}{\lambda d} \cdot d$$

   And it is the vanish point of the line.

   For every other line in the plane, we can derive the similar point, and all the vanishing point will form a line, which is the vanishing line of the plane.

(2) Assume that line L pass through the center of projection and the point $P(X, Y, Z)$ in the sphere. Then the equation of the line is $xX + yY + zZ - (X^2 + Y^2 + Z^2) = 0$. Plug it into the equation of the sphere $(x - X)^2 + (y - 0)^2 + (z - Z)^2 = r^2$, we can obtain

$$x = X * \frac{(X^2 + Y^2 + Z^2)}{(X^2 + Z^2 - r^2)}$$

$$y = 0$$

$$z = Z * \frac{(X^2 + Y^2 + Z^2)}{(X^2 + Z^2 - r^2)}$$

then we can get the square of the semi-major axis of the ellipse $a^2 = \frac{(X^2+Y^2+Z^2)}{(X^2+Z^2-r^2)}$ and the square of the semi-minor axis of the ellipse $b^2 = \frac{(X^2+Y^2+Z^2)}{(X^2+Z^2+r^2)}$, therefore, the eccentricity of the ellipse is

$$e = \sqrt{1 - \frac{a^2}{b^2}} = \frac{X}{\sqrt{X^2 + Z^- r^2}}$$

When the center of projection is outside the sphere, then the projection could be a parabola or hyperbola.

(3) Assume that $\delta Z = \delta Z_1 + \delta Z_2$, where δZ_1 is the depth error in front of P and δZ_2 is the depth error behind the P. Let $\frac{h}{Z} = tan\theta$, so that $\theta = atan\frac{h}{z}$, since the observer can visually discriminate angles to within $1^{'} = \frac{\pi}{10800}$, we have $\frac{h}{Z - \delta Z_1} = tan(\theta - \frac{\pi}{10800}) = tan(atan(\frac{h}{Z}) - \frac{\pi}{10800})$

Therefore,

$$\delta Z_1 = Z - \frac{h}{tan(atan(\frac{h}{Z}) - \frac{\pi}{10800})}$$

Similarly,

$$\delta Z_2 = Z - \frac{h}{tan(atan(\frac{h}{Z}) + \frac{\pi}{10800})}$$

Therefore,

$$\delta Z = 2Z - [\frac{h}{tan(atan(\frac{h}{Z}) - \frac{\pi}{10800})} + \frac{h}{tan(atan(\frac{h}{Z}) + \frac{\pi}{10800})}]$$

2. Visual Metrology: Infer Scene Geometry

   (1) slope $= \frac{H}{D}$

(2)

$Y_t = -V + tH, t = 0, 1, 2, ..., k$

$D_t = D_0 + (t-1)D, t = 0, 1, 2, ..., k$

(3)

From the figure, we could observe that

$$\frac{V - Y_t}{D_t} = \frac{y_t}{D_t - f}$$

therefore,

$$y_t = \frac{1}{D_t}(V - Y_t)(D_t - f), t = 0, 1, 2, ..., k$$

(4)

From (3), we can represent the $y_{t+1} - y_0$ with 3D scene points.

$$y_{t+1} - y_0 = \frac{1}{D_{t+1}}(V - Y_{t+1})(D_{t+1} - f) - \frac{1}{D_0}(V)(D_0 - f), t = 0, 1, 2, ..., k$$

(5)

From (3) we can represent the y_{k+1} as $y_{k+1} = \frac{1}{D_{k+1}}(V - Y_{k+1})(D_{k+1} - f)$

3. Training a Deep Neural Network

   (3)

   Cross-entropy loss function:

   When the number of class is 2, then the loss function is

$$L = -((y)logp + (1-y)log1-p)$$

   When there are more than 2 classes, says M, then the loss function is

$$L = -\sum_{c=1}^{M} y_{o,c} logp_{o,c}$$

   An interpret of the cross entropy function is that, it measures the performance of a classification model whose output is a probability value between 0 and 1. And cross-entropy loss increases as the predicted probability diverges from the actual label.

(4)

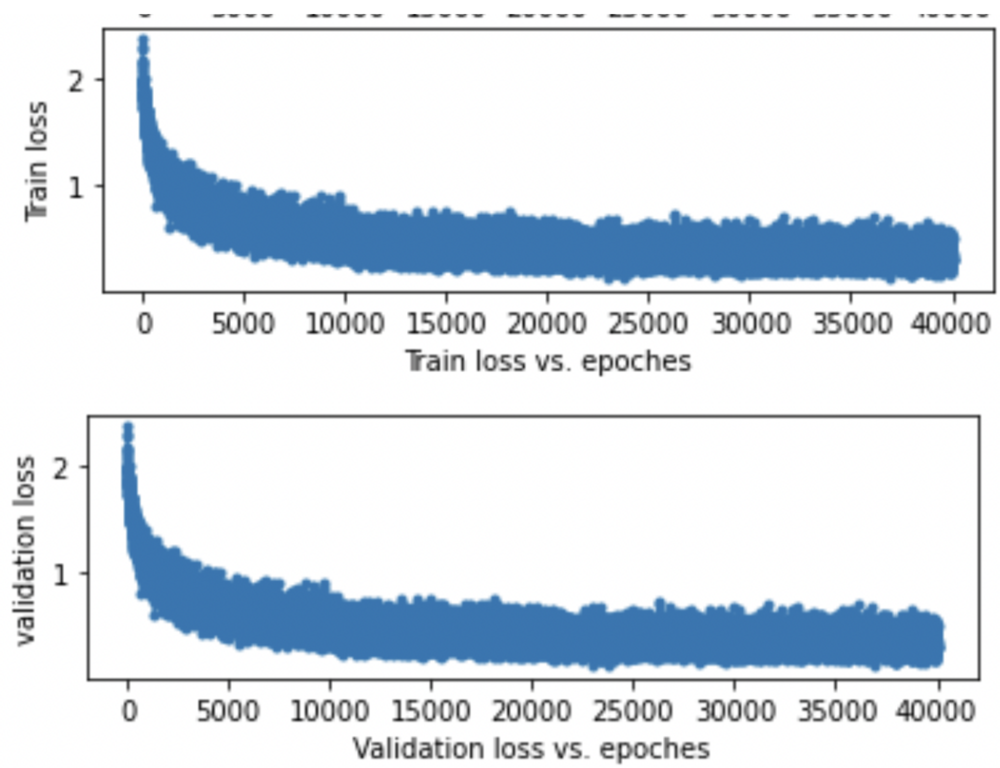

Figure 1: Loss value of training set and validation set

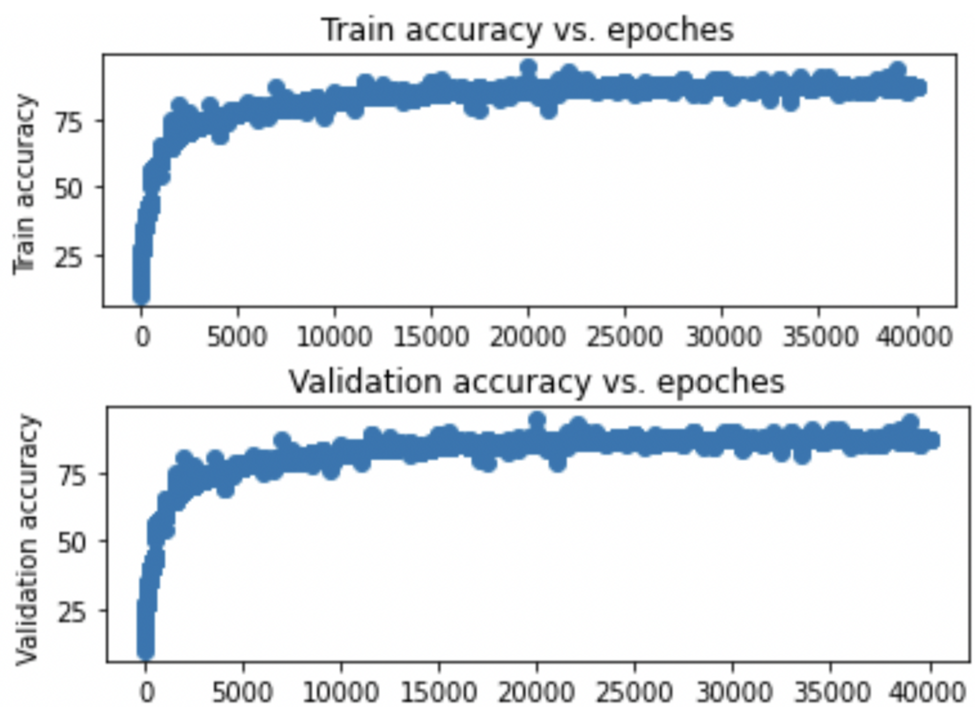

Figure 2: Accuracy of training set and validation set

Finally, the accuracy of the network is 85%.

4. Visualize Your Networks

(1)

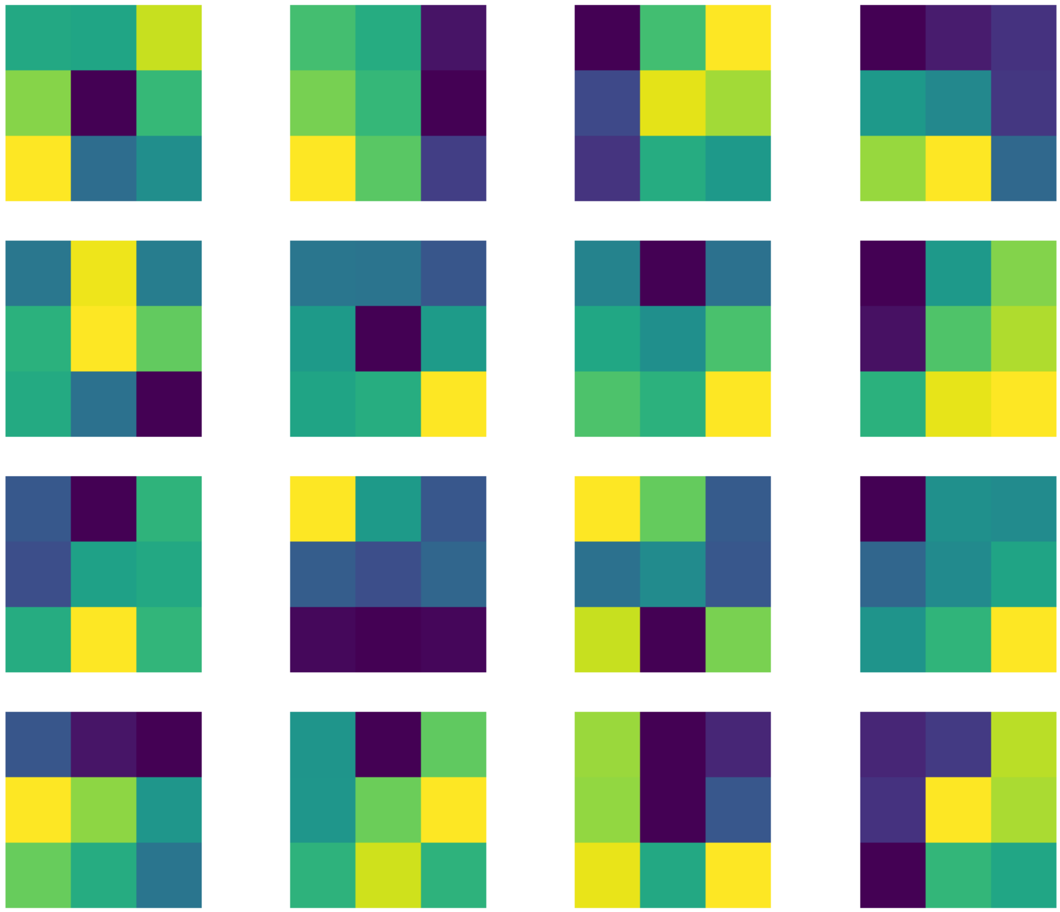

Figure 3: All the filters from the second layer

(2)

We can seen from figure3 that the first filter may extract the central feature of the original region because the central part has the highest value. What's more, the last filter may focus on the feature of the top left part.

(3)

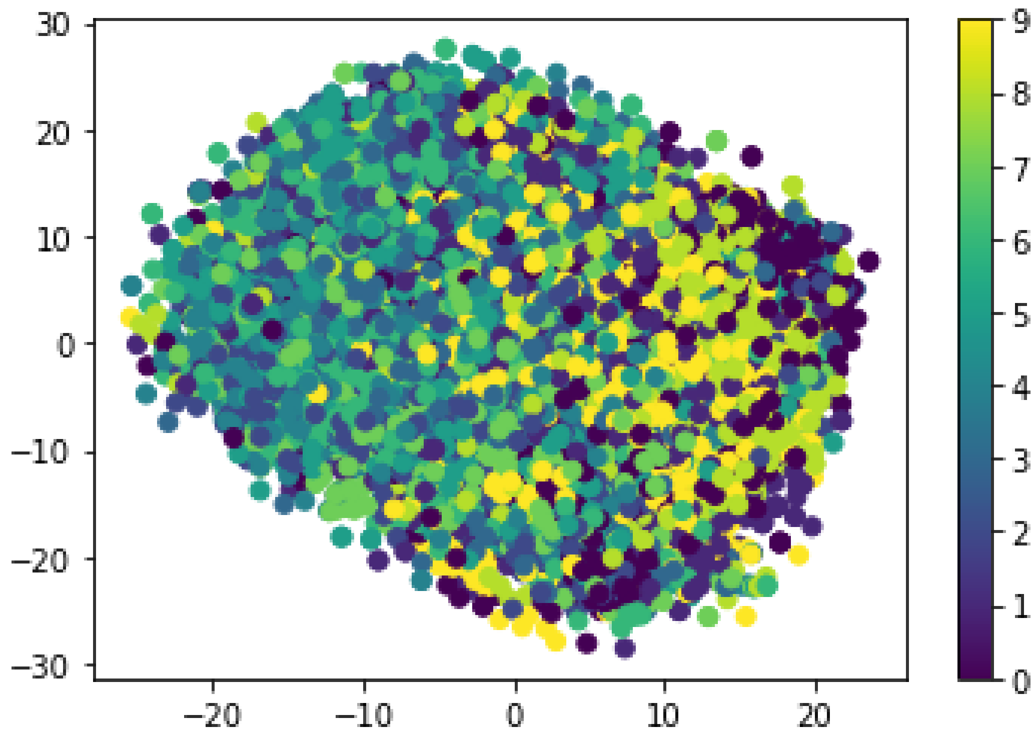

Figure 4: tSNE

From the picture, we can see that, the points that fall around (20, 10) is almost class 0, other classes could be observed to cluster in some ways. The tSNE can be used for dimension reduction, and illustrate the similarity in some way.