CS540 Introduction to Artificial Intelligence
Lecture 7

Young Wu
Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

June 30, 2021

Computer Vision Examples, Part I

Motivation

- Image segmentation
- Image retrieval
- Image colorization
- Image reconstruction
- Image super-resolution
- Image synthesis
- Image captioning

Computer Vision Examples, Part II

Motivation

- Style transfer
- Object tracking
- Visual question answering
- Human pose estimation
- Medical image analysis

Image Features

Motivation

- Using pixel intensities as the features assume pixels are independent of their neighbors. This is inappropriate for most of the computer vision tasks.
- Neighboring pixel intensities can be combined in various ways to create one feature that captures the information in the region around the pixel, for example, whether the pixel is on an edge, at a corner, or inside a blob.
- Linearly combining pixels in a rectangular region is called convolution.

Image Features Diagram

Motivation

One Dimensional Convolution

Definition

- The convolution of a vector $x = (x_1, x_2, ..., x_m)$ with a filter $w = (w_{-k}, w_{-k+1}, ...w_{k-1}, w_k)$ is:

$$a = (a_1, a_2, ..., a_m) = x * w$$

$$a_j = \sum_{t=-k}^{k} w_t x_{j-t}, j = 1, 2, ..., m$$

- w is also called a kernel (different from the kernel for SVMs).
- The elements that do not exist are assumed to be 0.

Two Dimensional Convolution

Definition

- The convolution of an $m \times m$ matrix X with a $(2k+1) \times (2k+1)$ filter W is:

$$A = X * W$$

$$A_{j,j'} = \sum_{s=-k}^{k} \sum_{t=-k}^{k} W_{s,t} X_{j-s,j'-t}, j, j' = 1, 2, ..., m$$

- The matrix W is indexed by (s, t) for $s = -k, -k+1, ..., k-1, k$ and $t = -k, -k+1, ..., k-1, k$.
- The elements that do not exist are assumed to be 0.

Convolution Diagram and Demo

Definition

Padding and Stride

Definition

- Unless specified otherwise, the pixels outside of the image are assumed to be 0. This is called zero padding.
- If there is no padding, then the dimension of the convolution will be smaller than the original image.
- Unless specified otherwise, the number of pixels to move the filters each time is 1. This is called a stride of 1.
- If the stride is equal to the filter size (length or width for a square filter), it is called non-overlapping convolution.

Image Gradient

Definition

- The gradient of an image is defined as the change in pixel intensity due to the change in the location of the pixel.

$$\frac{\partial I(s,t)}{\partial s} \approx \frac{I\left(s+\frac{\varepsilon}{2},t\right)-I\left(s-\frac{\varepsilon}{2},t\right)}{\varepsilon}, \varepsilon = 1$$

$$\frac{\partial I(s,t)}{\partial t} \approx \frac{I\left(s,t+\frac{\varepsilon}{2}\right)-I\left(s,t-\frac{\varepsilon}{2}\right)}{\varepsilon}, \varepsilon = 1$$

Image Derivative Filters

Definition

- The gradient can be computed using convolution with the following filters.

$$w_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}, w_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Sobel Filter

Definition

- The Sobel filters also are used to approximate the gradient of an image.

$$W_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, W_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Decomposition of Filters

Definition

- The Sobel filters can be decomposed into two one dimensional filters.

$$W_x = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}, W_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}$$

- It is significantly faster to do two one dimensional convolutions than to do one two-dimensional convolution.

Gradient of Images

Definition

- The gradient of an image I is $(\nabla_x I, \nabla_y I)$.

$$\nabla_x I = W_x * I, \nabla_y I = W_y * I$$

- The gradient magnitude is G and gradient direction Θ are the following.

$$G = \sqrt{\nabla_x^2 + \nabla_y^2}$$

$$\Theta = \arctan\left(\frac{\nabla_y}{\nabla_x}\right)$$

Gradient of Images Demo

Definition

Laplacian of Image

Definition

- The Laplacian of an image I is defined as the sum of the second derivatives.

$$\nabla^2 I(s,t) = \frac{\partial^2 I(s,t)}{\partial^2 s^2} + \frac{\partial^2 I(s,t)}{\partial^2 t^2}$$

$$\frac{\partial^2 I(s,t)}{\partial^2 s^2} \approx \frac{I(s+\varepsilon,t) - 2I(s,t) + I(s-\varepsilon,t)}{\varepsilon^2}, \varepsilon = 1$$

$$\frac{\partial^2 I(s,t)}{\partial^2 t^2} \approx \frac{I(s,t+\varepsilon) - 2I(s,t) + I(s,t-\varepsilon)}{\varepsilon^2}, \varepsilon = 1$$

Laplacian Filter

Definition

- The Laplacian can be computed using convolution with the following filters.

$$W_L = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\nabla^2 I = W_L * I$$

Edge Detection

Discussion

- Both the gradient and Laplacian of an image can be used to find edge pixels in an image.
- Images usually contain noise. The noises are not edges and are usually removed before computing the gradient.

2 Dimensional Gaussian Filter

Definition

- The Gaussian filter is used to blur images and remove noise in the image. A Gaussian filter with standard deviation σ is the following.

$$W_\sigma : (W_\sigma)_{s,t} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{s^2+t^2}{2\sigma^2}\right)$$

1 Dimensional Gaussian Filter

Definition

- The Gaussian filter can be decomposed into two one dimensional filters as well.

$$W_\sigma = w_\sigma * w_\sigma, (w_\sigma)_t = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

Gaussian Filter Example 3

Definition

- When filter size $k = 3$, and standard deviation $\sigma = 0.8$:

$$W_\sigma = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

- Sobel filter is approximately the combination of the gradient filter and the Gaussian filter.

Laplacian of Gaussian

Definition

- The Laplacian filter and the Gaussian filter are usually also combined into one filter called Laplacian of Gaussian filter (LoG filter).

$$W_{L,\sigma} : (W_{L,\sigma})_{s,t} = -\frac{1}{\pi\sigma^4}\left(1 - \frac{s^2+t^2}{2\sigma^2}\right)\exp\left(-\frac{s^2+t^2}{2\sigma^2}\right)$$

Difference of Gaussian

Definition

- The Laplacian of Gaussian filter is difficult to compute because it cannot be decomposed into two one dimensional filters. Therefore an approximation is used called the Difference of Gaussian filter (DoG filter).

$$W_{L,\sigma} \approx W_{\sigma} - W_{1.6\sigma}$$

LoG and DoG Diagram

Definition

Image Pyramids

Discussion

- There are edges at different scales of the image. Images are blurred and downsampled to get images with different scales.
- An image pyramid contains images at scales $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$

SIFT

Discussion

- Scale Invariant Feature Transform (SIFT) features are features that are invariant to changes in the location, scale, orientation, and lighting of the pixels.

Location and Scale Invariance

Discussion

- The gradient of pixels in a 16 by 16 region is used. The region is divided into 4 by 4 cells. Each cell contains the sum of the gradient in 8 different orientations (weighted by a Gaussian function).

$$x_j = \sum_{(s,t)\in \text{ cell } : \Theta(s,t) \in \left[\frac{\pi}{8} j, \frac{\pi}{8}(j+1)\right]} G(s,t)\, W_{0.5\cdot\sigma}(s,t)$$

for $j = 0, 1, ..., 7$

- This means each region is represented by a $4 \cdot 4 \cdot 8 = 128$ dimensional feature vector.

Histogram Binning Diagram

Discussion

Orientation Invariance

Discussion

- To make the features invariant to orientation, the dominant orientation in the region is usually calculated and the orientation of each pixel is rotated by the dominant orientation.

$$
\begin{aligned}
x_\theta &= \sum_{(s,t) \in \text{ cell } : \Theta(s,t) \in \left[\theta, \theta + \frac{\pi}{18}\right]} G(s,t)\, W_{1.5 \cdot \sigma}(s,t) \\
\text{for } \theta &= 0\frac{\pi}{18}, 1\frac{\pi}{18}, 2\frac{\pi}{18}, \ldots, 35\frac{\pi}{18} \\
\Theta^\star &= \arg\max_\theta x_\theta
\end{aligned}
$$

- Note that the dominant orientation is calculated using 36 bins, but the features are calculated using 8 bins. The Gaussian weights are calculated using different σ too.

Illumination and Contrast Invariance

Discussion

- To make the features invariant to different lighting, the 128-dimensional feature vectors are usually separately normalized (such that the sum is 1) and thresholded (values below 0.2 are made 0).

Keypoint Extraction

Discussion

- For computer vision tasks, SIFT feature vectors are calculated for a selected region around a small number of key points.
- The key points are local maxima and minima of the Laplacian of Gaussian of the image.

HOG

Discussion

- Histogram of Oriented Gradients features is similar to SIFT but does not use dominant orientations.
- 9 orientation bins are usually used for 8 by 8 cells. The gradient magnitudes are also not weighted by the Gaussian function.

$$x_j = \sum_{(s,t)\in \text{ cell } : \Theta(s,t) \in \left[\frac{\pi}{9} j, \frac{\pi}{9}(j+1)\right]} G(s,t), j = 0, 1, ..., 8$$

- The resulting bins are normalized within a block of 4 cells.

Classification

Discussion

- SIFT features are not often used in training classifiers and more often used to match the objects in multiple images.
- HOG features are usually computed for every cell in the image and used as features (in place of pixel intensities) in classification algorithms such as SVM.

Matching vs Classification Diagram

Discussion