The Austalian National University
School of Computing
Dylan Campbell

Semester 1, 2024
Tutorial 3

COMP/ENGN 4528/6528: Computer Vision

Question 1 Correlation and Convolution Calculations

1. If you were to apply the filter as a convolution:

$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 2 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

to the image

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 90 & 90 & 90 \end{bmatrix}$$

what would be the results? What would the results be if it was a correlation? Note that we are using the "valid" mode[1] in this question.

Solution.
Convolution
Flip the filter in both dimension (bottom to top, right to left) to obtain the convolution filter

$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 2 & -1 \\ 1 & 1 & 1 \end{bmatrix} \rightarrow_{\text{flip row}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & -1 \end{bmatrix} \rightarrow_{\text{flip column}} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 2 & 1 \\ -1 & -1 & 1 \end{bmatrix}$$

Then apply cross-correlation

$$\begin{bmatrix} 1 & 1 & 1 \\ -1 & 2 & 1 \\ -1 & -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 90 & 90 & 90 \end{bmatrix} = -90 - 90 + 90 = -90$$

Correlation

$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 2 & -1 \\ 1 & 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 90 & 90 & 90 \end{bmatrix} = 90 + 90 + 90 = 270$$

2. Consider the 5×5 image below. The pixel grey values are indicated by the values in the cells.

$$\begin{bmatrix} 3 & 2 & 1 & 2 & 4 \\ 2 & 1 & 3 & 200 & 3 \\ 6 & 7 & 8 & 7 & 9 \\ 8 & 100 & 6 & 6 & 7 \\ 7 & 9 & 6 & 8 & 8 \end{bmatrix}$$

   (a) Apply a 3×3 median filter to the image. Note that to avoid problems at the edges of the image you only need to calculate the filtered values for the central 3×3 region ("valid" mode).

[1] Read more about convolution mode from `https://au.mathworks.com/help/matlab/ref/conv2.html` (the "shape" parameter in the function documentation).

(b) Apply the vertical edge filter used by the Sobel edge detector to the image above. Again, you only need to calculate the results for the central 3×3 region ("valid" mode).

Note: in solving this problem, please use a "convolution" rather than "correlation" operator to implement the image filtering.

Solution.

(a) 3×3 median filter result

$$\begin{bmatrix} 3 & 3 & 4 \\ 6 & 7 & 7 \\ 7 & 7 & 7 \end{bmatrix}$$

Demonstrating how to calculate the top-left "3":
Find the corresponding region in the image:

$$\begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 6 & 7 & 8 \end{bmatrix}$$

Find the median of these 9 values

$$median(1, 1, 2, 2, 3, 3, 6, 7, 8) \rightarrow 3$$

(b) vertical Sobel filter result

$$\begin{bmatrix} 2 & 398 & 4 \\ 3 & 105 & 3 \\ -3 & -189 & 5 \end{bmatrix}$$

Demonstrating how to calculate the "398":
The vertical Sobel filter for convolution is

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

Since we are calculating the "convolution" operation, we flip the filter both horizontally and vertically:

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \rightarrow_{\text{flip row}} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \rightarrow_{\text{flip column}} \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

Find the corresponding region in the image:

$$\begin{bmatrix} 2 & 1 & 2 \\ 1 & 3 & 200 \\ 7 & 8 & 7 \end{bmatrix}$$

Then, we calculate the cross-correlation

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 2 & 1 & 2 \\ 1 & 3 & 200 \\ 7 & 8 & 7 \end{bmatrix} = -2 + 2 - 2 + 2 \times 200 - 7 + 7 = 398$$

Question 2 **Edge Detection**

1. The vertical and horizontal Sobel filters for correlation operations are given below

$$L_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \star L$$

$$L_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \star L$$

Please explain the effects of these two filters. How would you use these filters to obtain the orientation of the gradient?

Note: The $\star$ operator denotes the correlation operation.

Solution. The vertical Sobel filter detects vertical edges by computing the horizontal components of the image gradients. On the other hand, the horizontal Sobel filter computes the vertical components of image gradients. The direction of gradient in relation to x-axis can be calculated with $\arctan2(L_y, L_x)$.

Question 3 Correlation and Convolution Programming Questions

1. Complete the coding questions in `COMP4528_lab3_code.ipynb`.

   Solution. Not released, due to overlap with Assignment 1.

Question 4 Optional: Deeper Understanding to Convolution

1. **!!!** Compare the following two operations:

   (a) you first apply a 3×3 Gaussian filter (G) to an image I then followed by a Sobel filter (S);

   (b) you first apply a Sobel filter S to the Gaussian filter G, then convolve image I with this filter ($S * G$).

   Mathematically speaking, these two different operations should give you equivalent results, i.e., $S * (G * I) = (S * G) * I$, given that we are using the **"full" padding mode**[2] for convolutions (i.e., how convolutions are usually performed in signal processing). However, in practice one operation often produces better (more efficient) results than the other. Please explain why, and which one is your preferred edge detector?

   Solution. $(S * G) * I$ is better, because it can be more efficient.
   For two matrices with side lengths N and K, the full padding mode convolution will produce an output matrix with side length $(N + K - 1)$. So, the total number of calculation operations in full padding mode convolution is $(K \times K \times (N + K - 1)^2)$
   Assuming image has side length N,

   (a) Number of calculations when applying the filter G is: $(3 \times 3 \times (N + 3 - 1)^2) = 9(N + 2)^2$.
   Number of calculations when applying the filter S is: $(3 \times 3 \times (N + 2 + 3 - 1)^2) = 9(N + 4)^2$.
   Total number of calculations is: $9(N + 2)^2 + 9(N + 4)^2 = 18N^2 + 108N + 180$.

   (b) Number of calculations when applying filter S to filter G is: $(3 \times 3 \times (3 + 3 - 1)^2) = 225$, which produces a new 5×5 filter.
   Number of calculations when applying the new filter to the image is: $(5 \times 5 \times (N + 5 - 1)^2) = 25(N + 4)^2$
   Total number of calculations is: $25(N + 4)^2 + 225 = 25N^2 + 200N + 625$

[2] Read `https://brianmcfee.net/dstbook-site/content/ch03-convolution/Modes.html#full-mode` for more information about full padding mode in convolution

However, using the separability of S, G and $S * G$ filters,

(a) has $2 \times 3 \times (N+2)^2 + 2 \times 3 \times (N+4)^2 = 12N^2 + 72N + 120$ computations,

(b) has $2 \times 3 \times (3+3-1)^2 + 2 \times 5 \times (N+4)^2 = 10N^2 + 80N + 310$ computations.

So, when the image is large, option (b) will be more efficient.