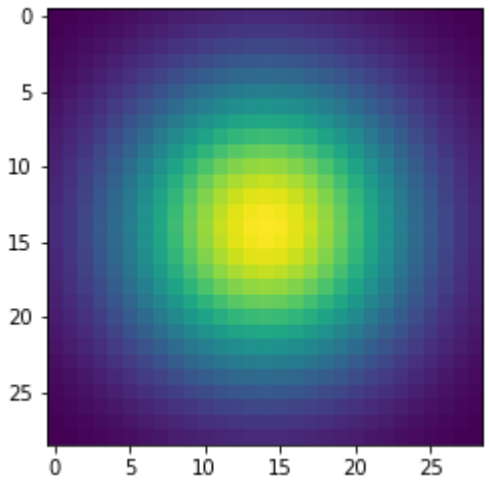
Computer Vision Project 1

刘嘉欣
211240076
211240076@smail.nju.edu.cn

Part 1: Image filtering

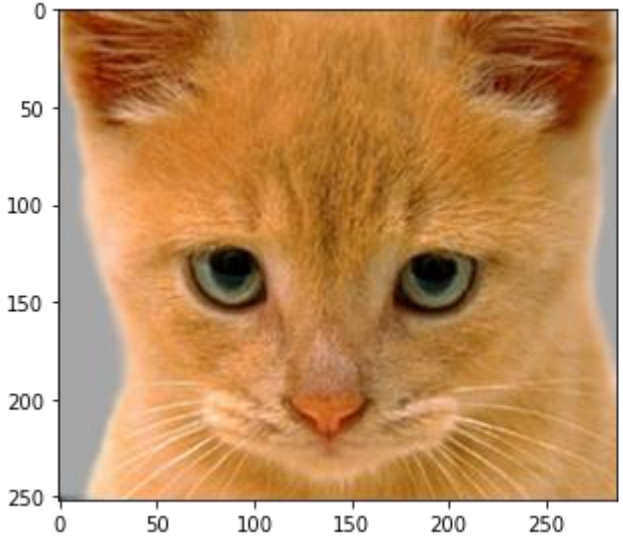


I apply padding to the image using numpy.pad which pads the image with k // 2 zeros before and after on axis 0, j // 2 zeros before and after on axis 1 and nothing on axis2, i.e. the second parameter is ((k // 2, k // 2),(j // 2, j // 2), (0, 0)).

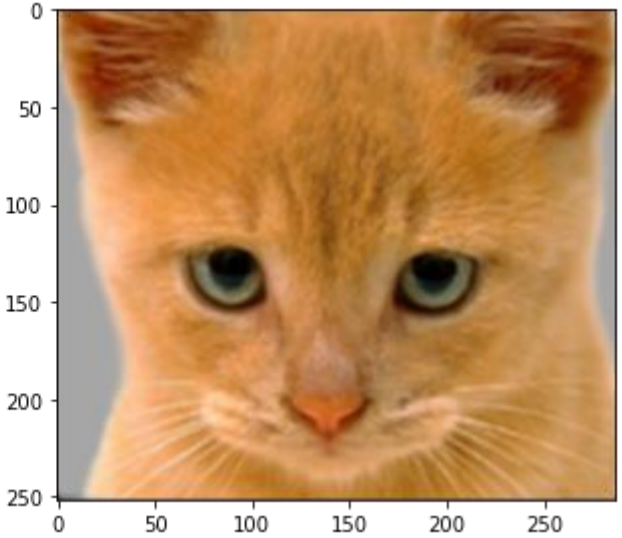
Then I compute the sum of the elementwise multiplication of filter and the corresponding sliced padded image using numpy.multiply() (*) and np.sum().

Part 1: Image filtering

Identity filter

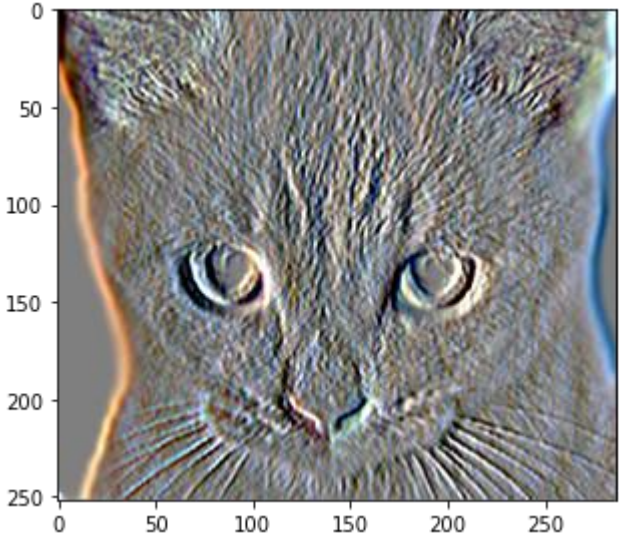


Small blur with a box filter

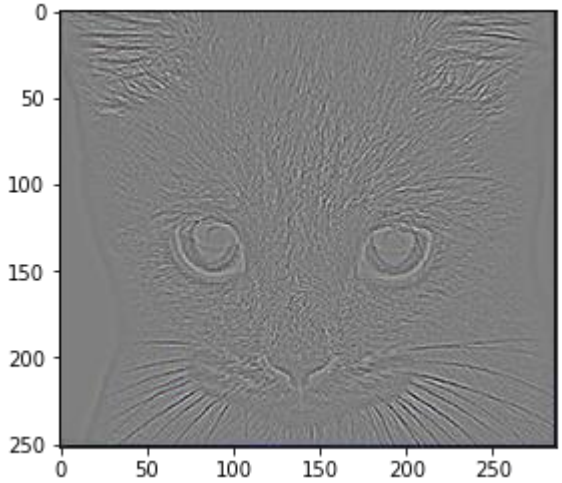


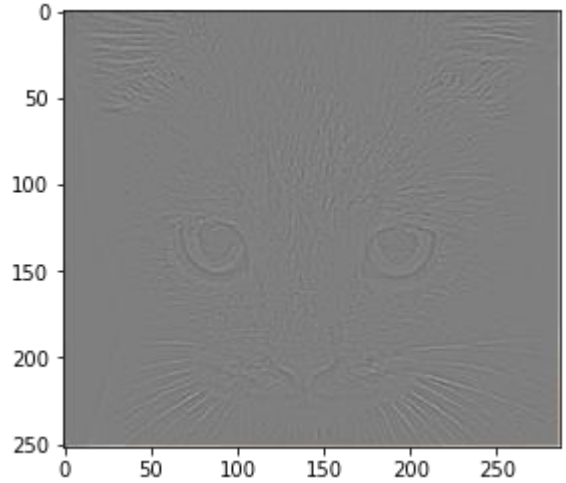
Part 1: Image filtering

Sobel filter



Discrete Laplacian filter





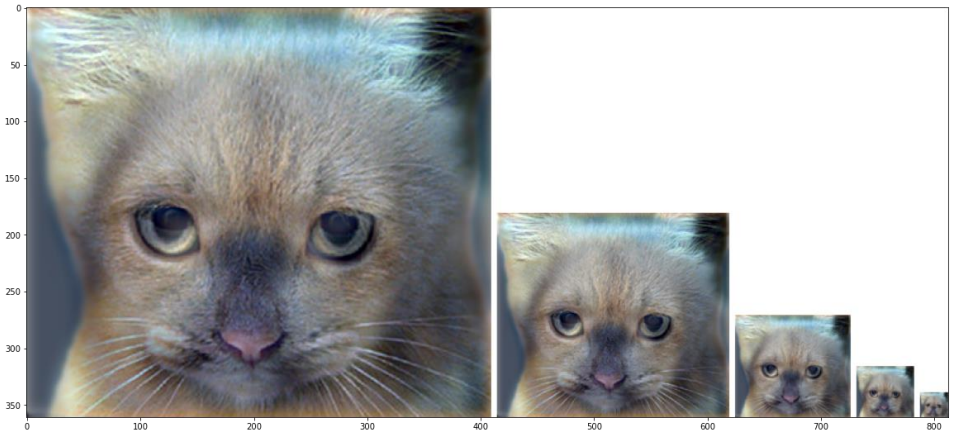
Part 1: Hybrid images

First we get low_frequencies by applying my_conv2d_numpy on the image1 and the filter.

Then high_frequencies of the image2 can simply get by subtracting its low-frequencies image.

In order to get the hybrid image, just plus them. Notice that we should ensure the output values are within the appropriate range [0, 1] by using numpy.clip(image, 0, 1)

Cat + Dog



Cutoff frequency: 5

Part 1: Hybrid images

Motorcycle + Bicycle

Cutoff frequency: 2

Plane + Bird

Cutoff frequency: 2

Part 1: Hybrid images

Einstein + Marilyn

Cutoff frequency: 3

Submarine + Fish

Cutoff frequency: 2

Part 2: Hybrid images with PyTorch

Cat + Dog

Motorcycle + Bicycle

Part 2: Hybrid images with PyTorch

Plane + Bird

Einstein + Marilyn

Part 2: Hybrid images with PyTorch

Submarine + Fish

Part 1 vs. Part 2

Part 2 is faster.

Part 3

[Consider a 1-channel 5x5 image and a 3x3 filter. What are the output dimensions of a convolution with the following parameters?
Stride = 1, padding = 0?
Stride = 2, padding = 0?
Stride = 1, padding = 1?
Stride = 2, padding = 1?]

Formula: size' = (size-k+2*padding)/stride+1, so the output dimensions of a convolution with the following parameters are:

(3, 3), (2, 2), (5, 5), (3, 3).

[What are the input & output dimensions of the convolutions of the dog image and a 3x3 filter with the following parameters:
Stride = 1, padding = 0
Stride = 2, padding = 0
Stride = 1, padding = 1
Stride = 2, padding = 1?]

(3, 361, 410) -> (1, 12, 359, 408)
(3, 361, 410) -> (1, 12, 180, 204)
(3, 361, 410) -> (1, 12, 361, 410)
(3, 361, 410) -> (1, 12, 181, 205)

Part 3

[How many filters did we apply to the dog image?]

4 filters.

[Why do the output dimensions adhere to the equations given in the instructions handout?]

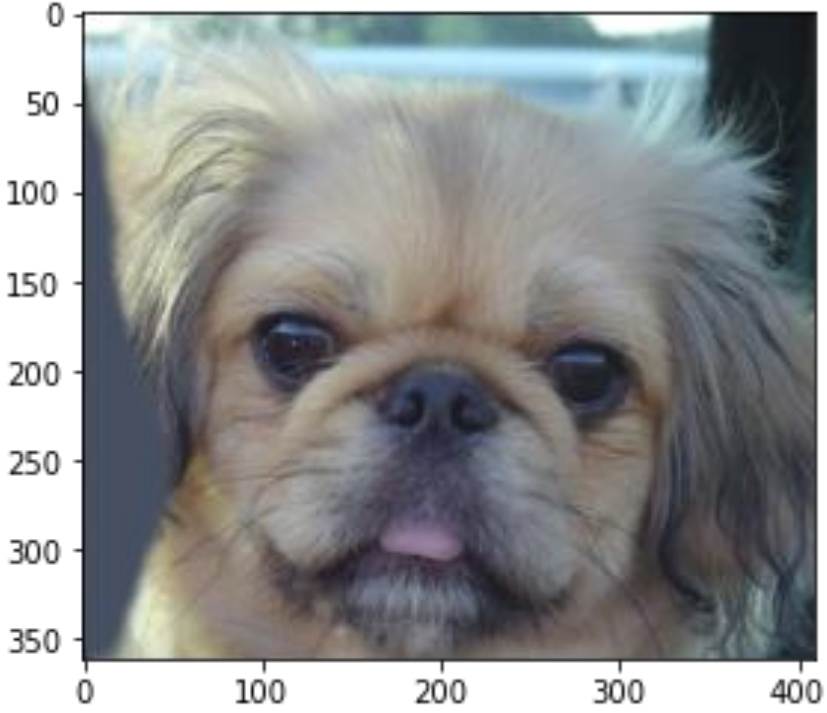
Clearly, the number of output channels must be equal to the 1st parameter of the kernel.
Formula: N=(W-F+2*P)/S+1
Padding the image will enlarge the image size by 2 * padding for height and width, while stride will decrease the output size by portion since it skips data.

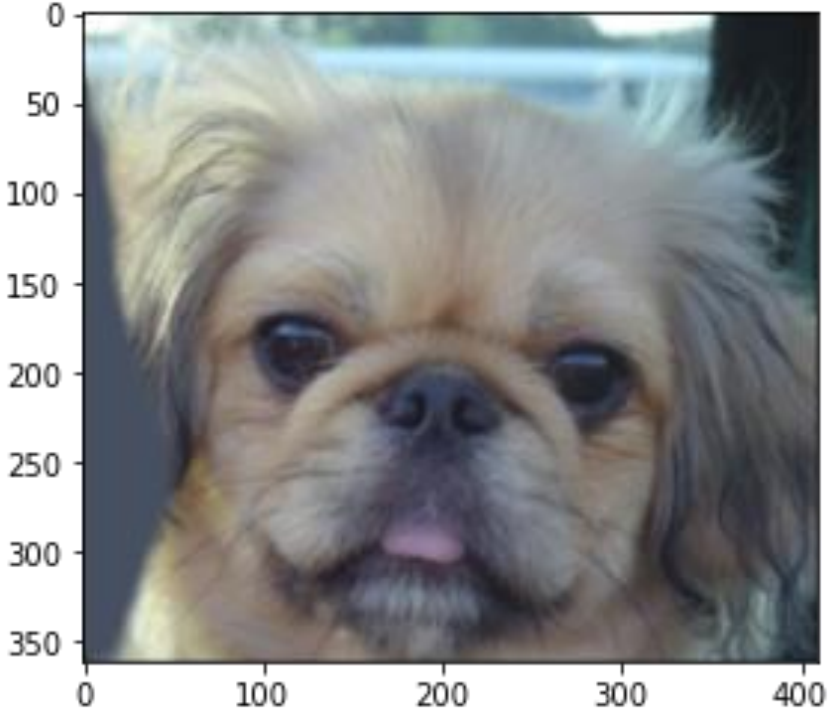
Part 3

[What is the intuition behind this equation?]

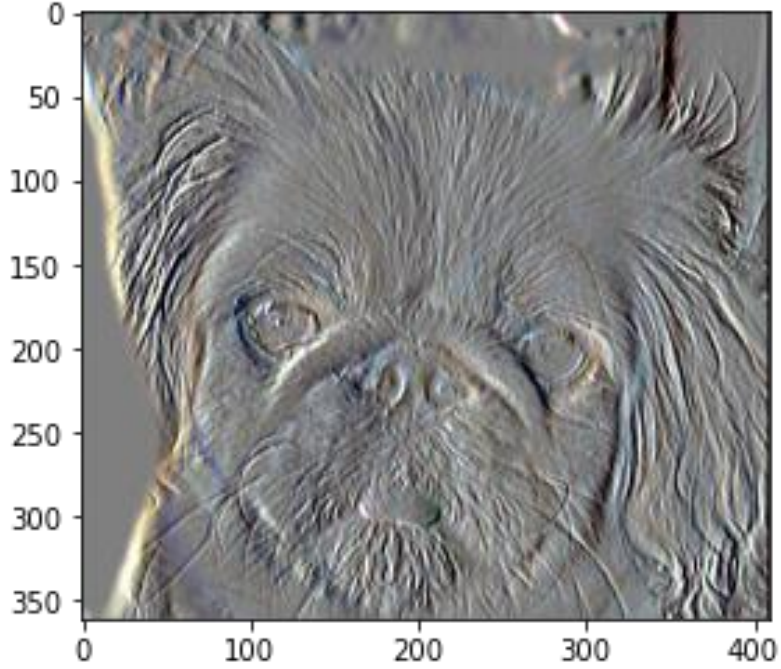
The equation for calculating the output dimensions of a convolutional layer takes into account the size of the filters, the amount of padding, and the stride. The filter size determines the receptive field, padding preserves spatial dimensions, and stride controls the step size of the filter. By subtracting the filter size from the input dimensions, adding padding, dividing by the stride, and adding 1, we determine the number of times the filter can be applied and thus the output size. Adjusting these factors allows for control over the spatial dimensions and capacity of the model.

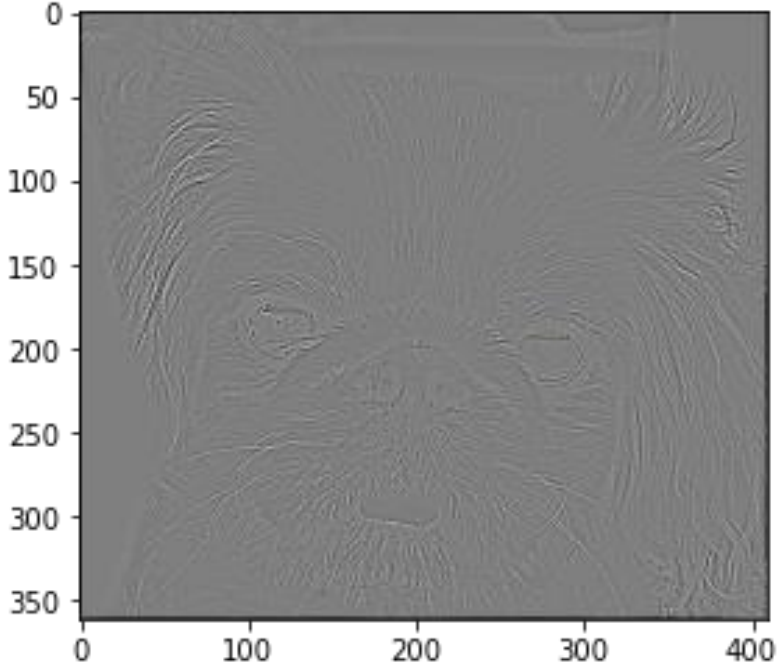
Part 3


0
50
100
150
200
250
300
350
0
100
200
300
400


0
50
100
150
200
250
300
350
0
100
200
300
400

Part 3


0
50
100
150
200
250
300
350
0
100
200
300
400


0
50
100
150
200
250
300
350
0
100
200
300
400

Conclusion

[How does varying the cutoff frequency value or swapping images within a pair influences the resulting hybrid image?]

varying the cutoff frequency value:
Varying the cutoff frequencies in a hybrid image influences the balance between low-frequency structure and high-
frequency details. Higher cutoff frequencies result in more prominence of high-frequency details, while lower cutoff frequencies emphasize low-frequency structure.

swapping images within a pair:
Swapping images within a pair when creating a hybrid image changes the dominant fe atures. Swapping the low-frequency image emphasizes low-
frequency structure, while swapping the high-frequency image highlights high-frequency details.