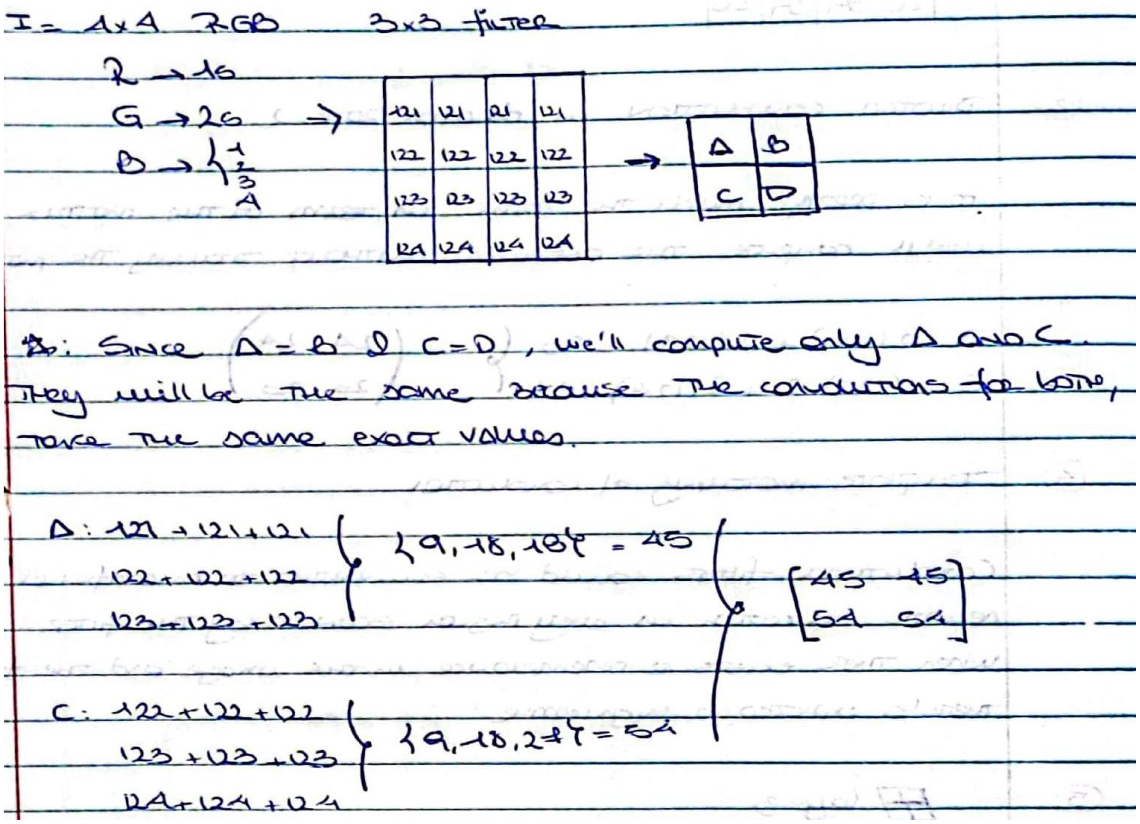
CS-577 Deep Learning. Assignment 2

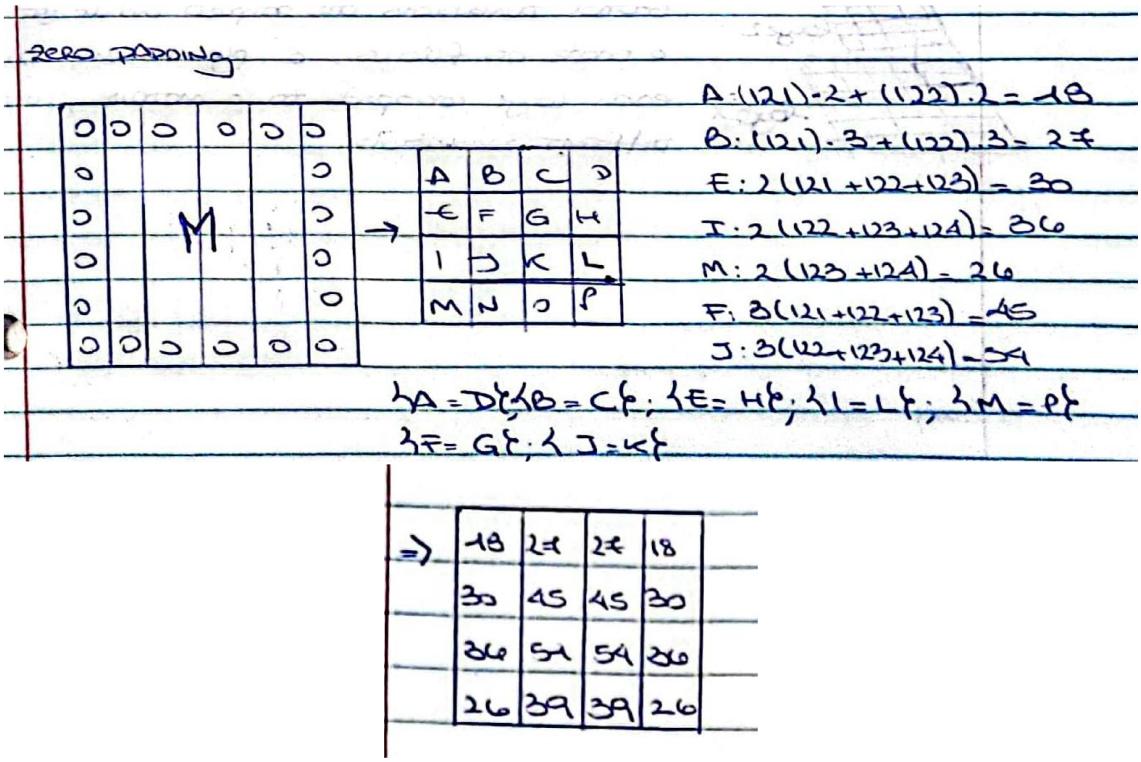
Guillermo Lopez-Areal

Illinois Institute of Technology

1. Let I be a 4×4 RGB image where the R channel is all 1-s and G channel is all 2-s. The B channel has a value of 1 in its first row, a value of 2 in its second row, a value of 3 in its third row, and a value of 4 in its 4th row. Compute the convolution of this image with a 3×3filter having all ones without zero padding.



2. Repeat the previous question with zero padding.



3. Repeat the previous question when using dilated (atrous) convolution with a dilation rate of 2.

DILATED CONVOLUTION W. DILATION RATE 2

IT IS PRETTY MUCH THE SAME IN TERMS OF THE MATRIX, SO WE'LL COMPUTE THE ANSWER WITHOUT USING THE MATRIXES

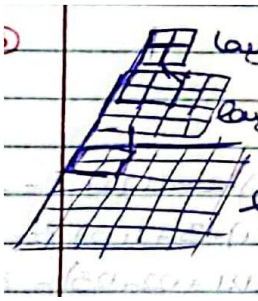
$A: 2(121) + 2(123) = 20$
$E: (122) \cdot 2 + 2(124) = 24$
$\Rightarrow \begin{pmatrix} 24 & 24 \\ 20 & 20 \end{pmatrix}$

4. Explain the template matching interpretation of convolution.

TEMPLATE MATCHING OF CONVOLUTION

CONVOLUTIONAL FILTERS COULD BE EQUIVALENT TO AN ARTIFICIAL NEURONS ACTIVATION IN EVERY REGION SCANNED BY THE FILTER. WHEN THERE EXISTS A RESEMBLANCE IN THE IMAGE AND THE FILTER, THERE'S EXPECTED A HIGH RESPONSE

5. Explain how multiple scale analysis can be achieved with a fixed window size (using a pyramid).



SPATIAL DIMENSIONS ARE SAMPLED AND WE GET A SHAPE AS FOLLOWS: A PYRAMID SHAPE WHERE EACH LAYER CORRESPONDS TO A MATRIX, WITH DIFFERENT RESOLUTIONS

6. Explain how to compensate for spatial resolution decrease using depth (number of channels) and the purpose for doing so.

AS SPATIAL DIMENSIONS DECREASE, DEPTH INCREASE TO COMPENSATE FOR REDUCED COEFFICIENTS. THIS IS DONE TO HELP, TO FACILITATE THE LEARNING OF DISTINCT AND COMPLEX STRUCTURES ARE MADE THEM SMALLER TO FIT THEM IN THE NETWORK

7. Given a $128 \times 128 \times 32$ tensor and 16 convolution filters of size $3 \times 3 \times 32$, what with be the size of the resulting tensor when convolving without zero padding.

$\text{SIZE} = \frac{W - K + 2P}{S} + 1 \Rightarrow 126 \times 126 \times 16$

8. Repeat the previous question when using a stride of 2.

$63 \times 63 \times 16$

9. Explain how the number of channels can be reduced using a 1×1 convolution.

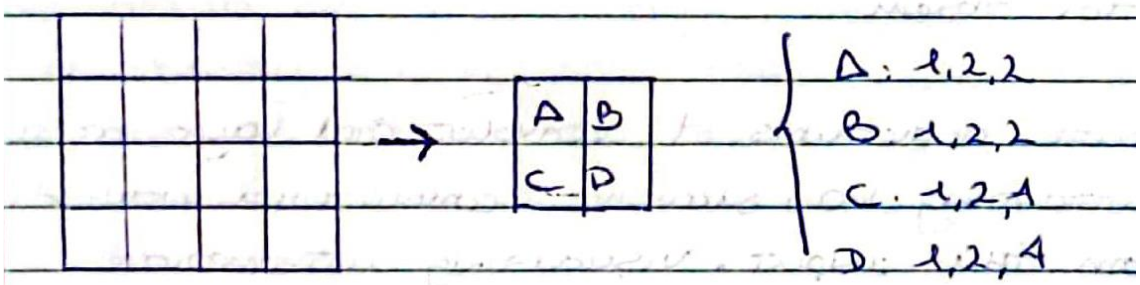
convolution with N 1x1 filter results N number of channels output. With a fewer n.of filters it can also be reduced

10. Explain the interpretation of convolution layers and the difference between early and deeper convolution layers.

used to extract image patterns, vectorize image windows and filter extending full length of image

11. Let I be an image as in question 1. Write the result obtained using max pooling with a 2×2 filter with a stride of 2.

Pooling with 2x2 filter with stride 2



12. Explain the purpose of pooling.

Pooling is made for down sampling dimensions (depth unchanged).

13. Explain the purpose of data augmentation and when it is most useful.

Data augmentation: augment data for better generalization using modifications of available data. It is commonly used when there are small datasets

14. Explain the purpose of transfer learning and when it is most useful.

Transfer learning is the use of pre-trained model's weights. It is commonly used to get a better performance when training new models, specially in convolutional neural networks for image recognition

15. Explain the need for freezing the coefficients of the pre-trained network.

Freezing is done to not push the gradients through the pretrained base when starting to train the new classifier on the top of the network.

16. Explain how he coefficients of a pre-trained network can be fine-tuned.

Fine tuning is done after training the classifier, by unfreezing some layers on top of the conv-base and retraining to let the model to fit the data

17. Explain the purpose of inception blocks.

Inception blocks allow to use multiple filter sizes in a single image block in convolutional layers

18. Explain the advantage of residual blocks.

Residual blocks help to avoid gradient vanishing in deep convolutional neural network.
Zero weights in the block produce identity instead of destroying the signal
Residual blocks create an identity mapping to activations earlier in the network, in order to thwart the performance degradation problem.

19. Explain how intermediate activations of convolution layers can be visualized given an input. What is the purpose for doing so?

Intermediate activations of convolutional layers are useful for understanding how successive convolutional network layers transform their input. Visualizing intermediate activations consists of displaying the feature maps that are outputs by various convolution layers in a network. This gives a view into how an input is decomposed into the different filters learned by the network

20. Explain how the filter weight of the trained convolution layers can be visualized (using gradient ascent to find the input with maximal response). What is the purpose for doing so?

The way in which the filter weights, and to determine the most efficient outcome, is by performing gradient descent, to do it can find the input features, from where the outcome of the filters is maximized. The purpose of doing so, is so we can determine which filter to use, and how such filter looks are · each epoch

21. Explain how the heatmap of class activation can be visualized for a specific image and class. explain how pooled gradients can be used to weight channels in this visualization. Explain the purpose of this visualization.

After sending a certain image to our neural network, and computing the we compute gradient descent with respect to each channel. And what it's interesting, is that the neural network decides which CNN channel is more crucial for predicting the results. And this idea is done by computing the gradients in each channel and comparing them to others, so as the one that is more significant, will receive a higher weight. Similarly those who are less significant will receive a lower weight. And finally, we perform global average pooling, so as we add up and combine every possible channel and, we have finally our output. This is done, so as we can visualize which one of the image helped the most the neural network to determine to which class a certain image belongs.