AIST4010: Foundation of Applied Deep Learning
Lecture 8 Scribing: CNN++

Lecturer: Professor Yu LI Scriber: Man Ho LAM (1155159171)

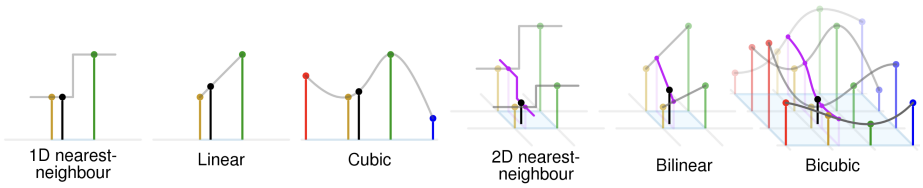
4 February 2024

1 Image Super-Resolution

Given a small size (low resolution) image, we want to predict a super-resolution image by filling the missing values to construct a higher dimension and resolution image.

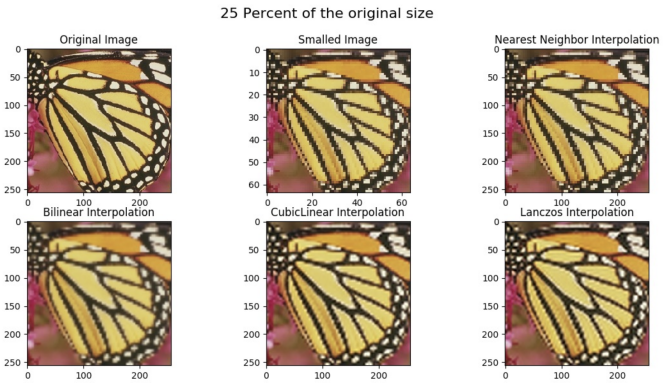
1.1 Interpolation

Interpolation is a traditional implementation for the image super-resolution algorithm. Assuming the image is smooth, it uses different functions to predict and fill in the missing values based on the around pixels with well-known values.



However, interpolation has following shortcomings:

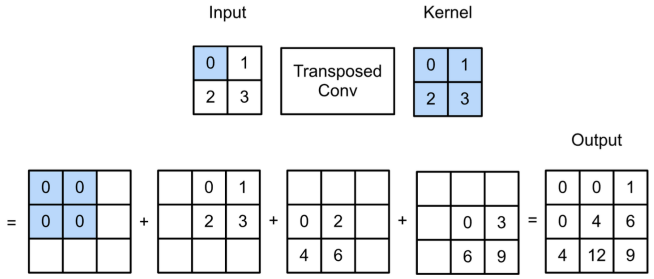
- It completely ignores the data
- The results have a blurring effect
- It will amplify the noise within the super-resolution images
- It ignores the semantic meaning and the far-away information of the images



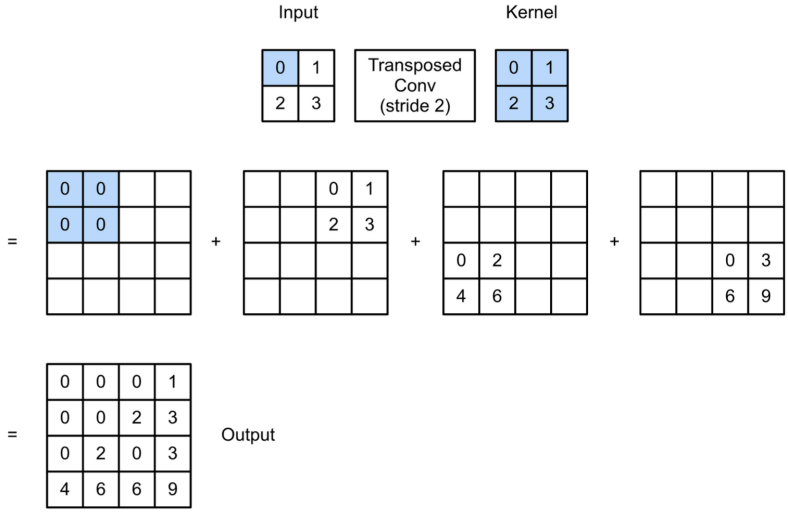
Ideally, we hope the algorithm itself can understand the problem, and interpolate the values based on the problem, instead of some pre-fixed functions.

1.2 Transposed Convolution

Given a learnable kernel, which is determined by the data and the training algorithm. Multiply every pixel in the input with the kernel to the corresponding pixels of the feature map with a default 0 value for the other pixel. Construct the output feature map by summing up all the feature maps for all pixels in the input.



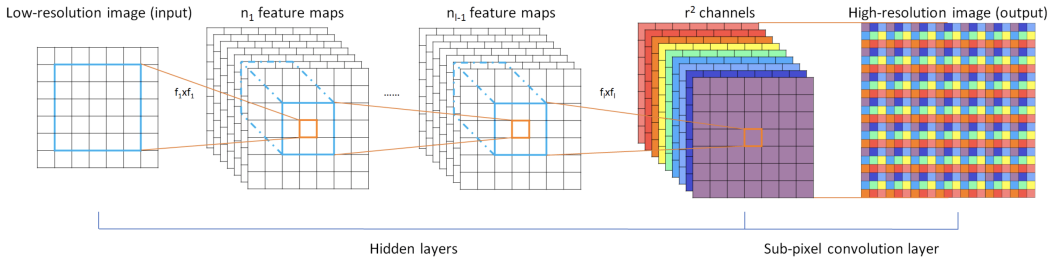
Transposed Convolution with stride = 1.



Transposed Convolution with stride = 2.

1.3 Pixel Shuffler: Sub-pixel Convolution

Sub-pixel Convolution is used to rearrange the dimension of the output image after the hidden layers. Given that there are r^2 channels with each having $f_l \times f_l$ pixels (dimension). In the sub-pixel convolution layer of pixel shuffler, it re-shuffles all the pixels to a one channel feature map. The dimension of output feature map is $(f_l \cdot r) \times (f_l \cdot r)$.



2 Image Segmentation

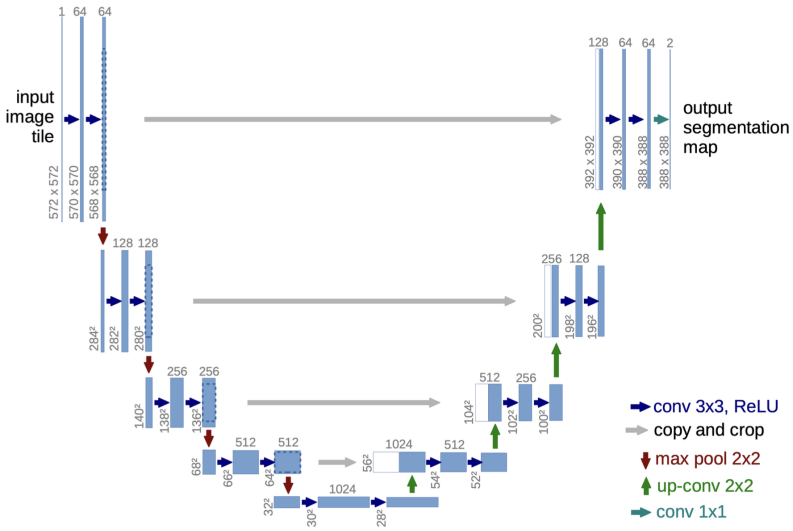
In this task, we want to divides the image into several constituent regions. There are two different segmentation methods:

- **Semantic segmentation**: label each pixel to a class
- **Instance segmentation**: label each pixel to an instance

The difference is if given an image with two dogs, the semantic segmentation will assign the pixels belonging to two dogs as "dog", while the instance segmentation also needs to mention "dog 1" and "dog 2".

2.1 UNet

The input image will first go through the convolution operations to gradually decrease the feature map dimension. Then, the dimension will increase further, so that the output dimension is the same as the input dimension. There are some shortcut point-wise additions within the network.



Question: Why reduce the dimension?
Answer: Because we want to extract some semantic information and features from the image.
Question: Why shortcut?
Answer: Because for those local regions, we want some local information, which is preserved in the original image.

URelpy Section

Question: Transpose convolutions are equivalent to up-sampling the input maps by inserting zeros, and then performing a conventional convolution operation, true or false?
Answer: True, because we add the values of the transposed convolution from different pixels into the final value in the output feature map is equivalent to running a traditional convolution and getting the result in the output feature map.

3 Applications of Convolutional Neural Network

Where can we apply CNN?
We want to take advantage of the properties within the image, translation invariance and locality. Therefore, as long as they have some local special patterns, with spatial/temporal correlation (local information), it is possible to apply CNN.

3.1 1D pattern

- **Text:** Since natural language contains patterns, it is possible to do the sentiment sentence classification.
- **Bio-sequences:** Those sequences consist of special information, which can use CNN to make some predictions.
- **EEG data:** We believe that there are some patterns of some special information in EEG.

3.2 2D pattern

- **Images classification**
- **Object detection**
- **Image segmentation**
- **Image generation**
- **Useful image datasets:** MNIST, CIFAR-10/100, Fashion-MNIST, CelebA, ImageNet.

3.3 3D pattern

- **CT scan**
- **Molecule**
- **Video**
- **3D object detection**

URelpy Section

Question: Which of the following data, CNN may not have a significant advantage over the fully-connected network?
A. Image B. Video C. Tabular D. Text E. Bio-sequences
Answer: C. Tabular, because there is no special information, and the features may be independent of each other.