CSE3042
Machine Intelligence for Medical Image Analysis

U-Net Model

Convolutional Networks for Biomedical Image Segmentation

Yash Bhootda
20BAI1060

Masterplan

1. Before U-Net
2. Introduction
3. Architecture
4. Why Medical Image Analysis
5. Analogy
6. Working with code
7. References

Before U-Net

Sliding Window Convolution Network

A sliding window convolutional network is a type of neural network architecture that uses a small "window" to slide across an input image or data and applies convolutional operations to the local region covered by the window, producing a set of feature maps that capture local patterns and structures in the input. It used the method of setting up the network of sliding window architecture by making the class label of each pixel as separate units by providing a local region (patch) around that pixe

Limitations

1. Due to considering each pixel separately, the resulting patches had a lot of overlap, leading to redundancy.
2. The training procedure was slow and resource-intensive.

Introduction

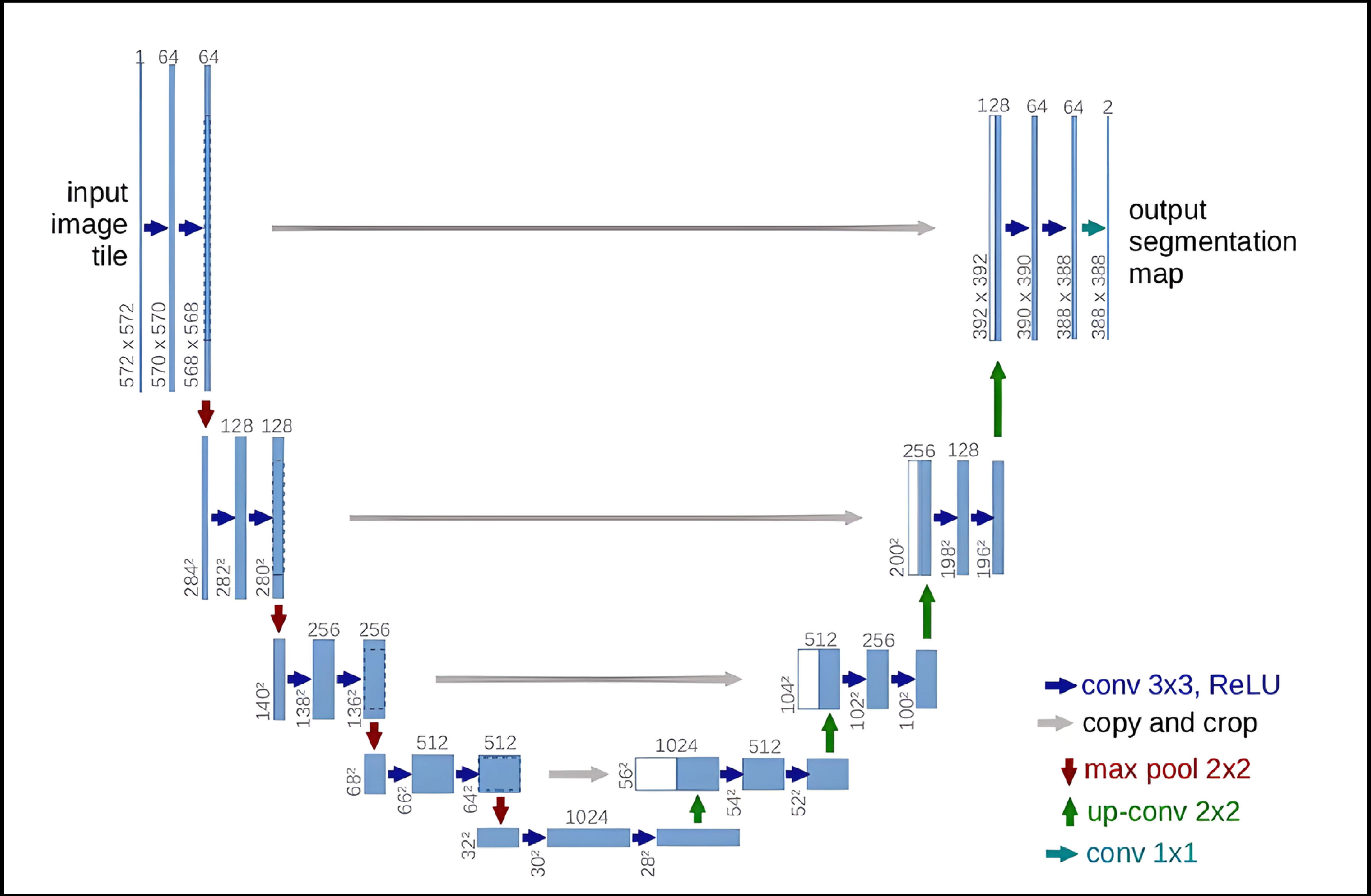
- The U-Net model is a type of convolutional neural network architecture that was first proposed for biomedical image segmentation tasks. It is named after its U-shaped architecture.

- The encoder path of the U-Net model typically consists of a series of convolutional and pooling layers that downsample the input image, allowing the network to capture increasingly abstract and high-level features. The decoder path, on the other hand, uses upsampling and concatenation with the corresponding feature maps from the encoder path to reconstruct the segmentation map at the original resolution.

- One of the key innovations of the U-Net model is the use of skip connections between corresponding layers in the encoder and decoder paths. These skip connections allow the network to retain information from earlier stages of processing and combine it with information from later stages

- The input size and the output size are the same

Architecture



Why Medical Image Analysis

- **High spatial resolution**: Medical images often have a very high spatial resolution ie a lot of pixels , which can make it challenging to process the entire image at once.
- encoder-decoder architecture , downsampling in the encoder path and upsampling in the decoder path.

- **Limited training data**: Medical image datasets are often small due to the difficulty and expense of obtaining medical images.
- skip connections which enable the network to learn more robust features from the available data.

- **Complex image structures**: Medical images can have complex structures such as overlapping organs or tissue layers.
- its multi-scale feature extraction capabilities, which allow it to capture features at different levels of abstraction.

Analogy

- Imagine you are building a puzzle of a beautiful landscape picture. However, the puzzle is very large and has many small pieces, making it difficult to see the overall picture. To help you solve the puzzle, you decide to use a two-part approach.

- First, you sort the puzzle pieces into different groups based on their color and shape, making it easier to see the different features of the landscape. This is similar to the encoder network of the U-Net model.

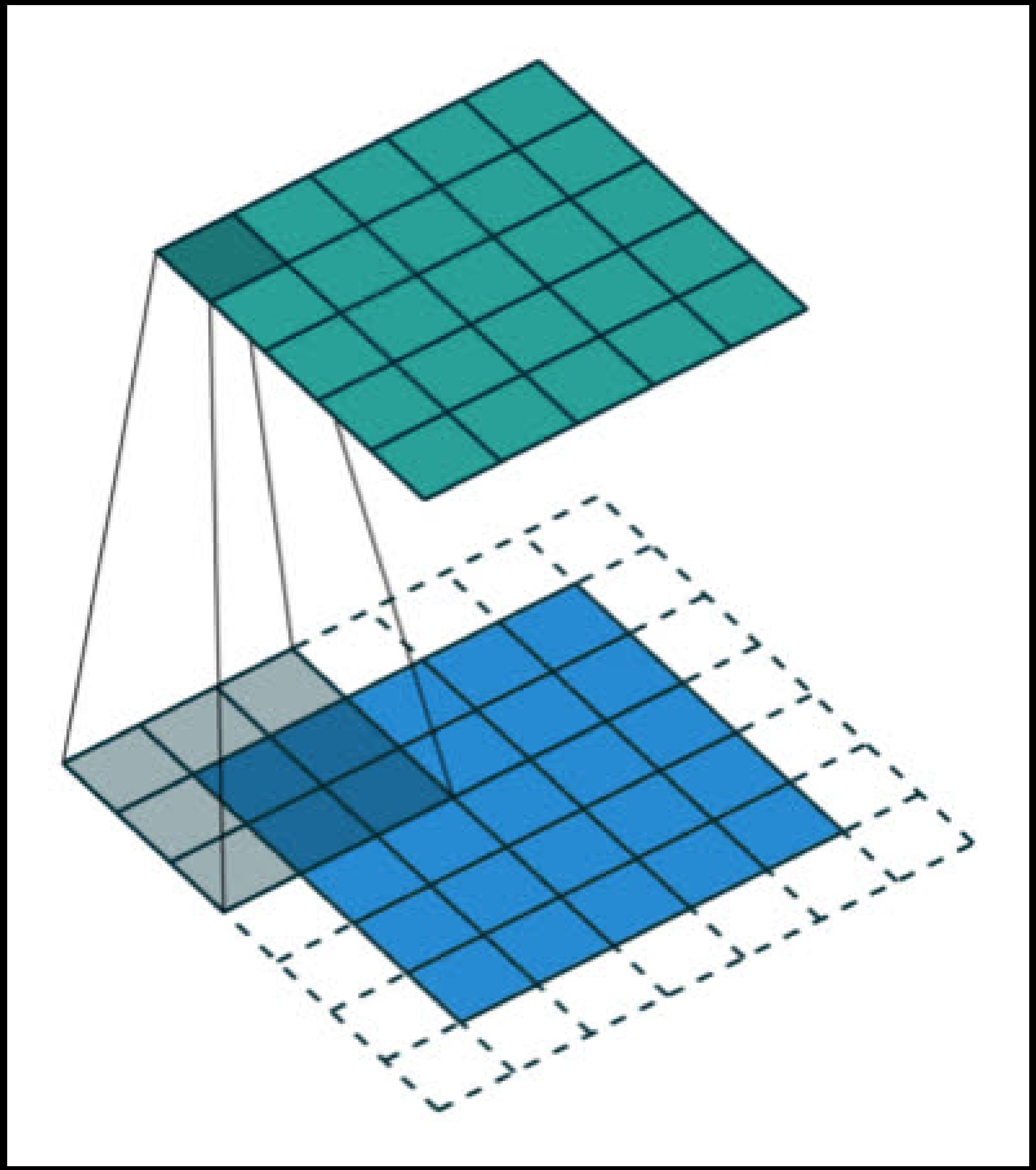
- Next, you start piecing together the puzzle by focusing on one group of pieces at a time, gradually building up the landscape from its individual parts. However, you also keep referring back to the completed sections of the puzzle to ensure that the new pieces fit together seamlessly. This is similar to the decoder network of the U-Net model, which upsamples the feature maps and reconstructs the segmentation map, while also incorporating information from the encoder network through skip connections.

- In this way, the U-Net model is like solving a puzzle, where the encoder network helps to break down the input image into smaller, more manageable parts, and the decoder network uses these parts to reconstruct a larger, more complete picture.

Working with code - Downscaling

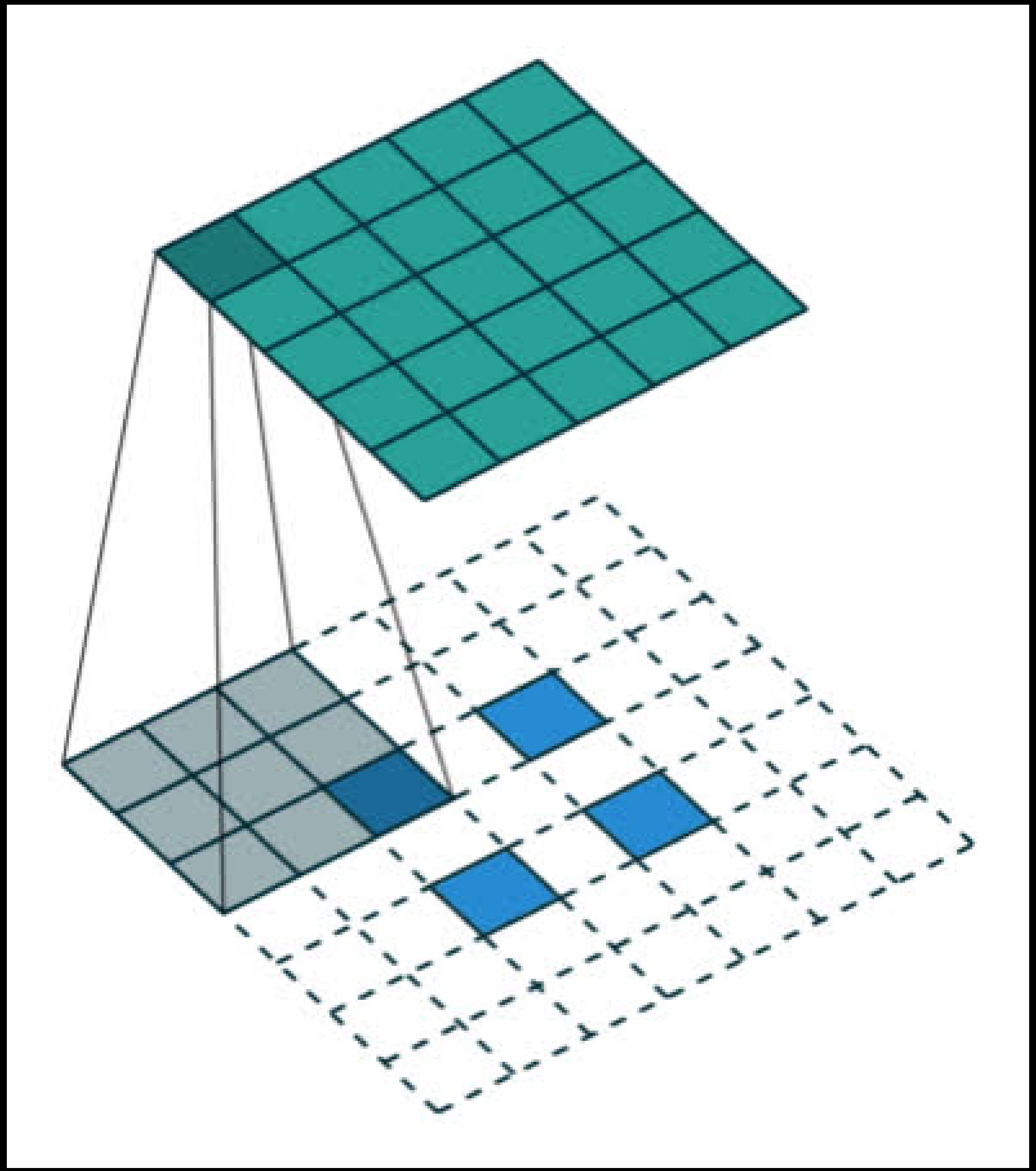
- **Notice that each process constitutes two convolutional layers**, and the number of channel changes from 1 → 64, as convolution process will increase the depth of the image. The red arrow pointing down is the max pooling process

- The last layer does not have a max pooling layer. The image at this moment has been resized to 16x16x1024. Now let's get to the expansive path.



Working with code - Upscaling

- Transposed convolution is an upsampling technic that expands the size of images.
- After the transposed convolution, the image is upsized from 16x16x1024 → 32x32x512, and then, this image is concatenated with the corresponding image from the contracting path and together makes an image of size 32x32x1024.



MCQ

What is the purpose of the skip connections in the U-Net architecture?

a) To reduce overfitting in the model
b) To concatenate the feature maps from the encoder and decoder
c) To increase the depth of the model
d) To speed up the training process

MCQ

What is the unique feature of the U-Net model?

a) Skip Connection

b) Keeps Input and Output size equal

c) It is U in shape.

d) Both A and B

MCQ

What is the role of the max-pooling operation in the U-Net architecture?

a) To reduce the size of the feature maps

b) To increase the depth of the model

c) To compute the loss function

d) To regularize the weights of the model

References

1. https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5
2. https://towardsdatascience.com/types-of-convolutions-in-deep-learning-717013397f4d
3. https://github.com/yash-bhootda/U-Net (Code , ppt and architecture)
4. https://www.kaggle.com/code/phoenigs/u-net-dropout-augmentation-stratification/notebook
5. https://chat.openai.com/chat

Thank You

(For pretending to listen)