Intro to Neural Nets

Session 6: Advanced CNNs

Session Agenda

More Modern Image-Network Architectures

- Mitigating vanishing gradients in deep networks.

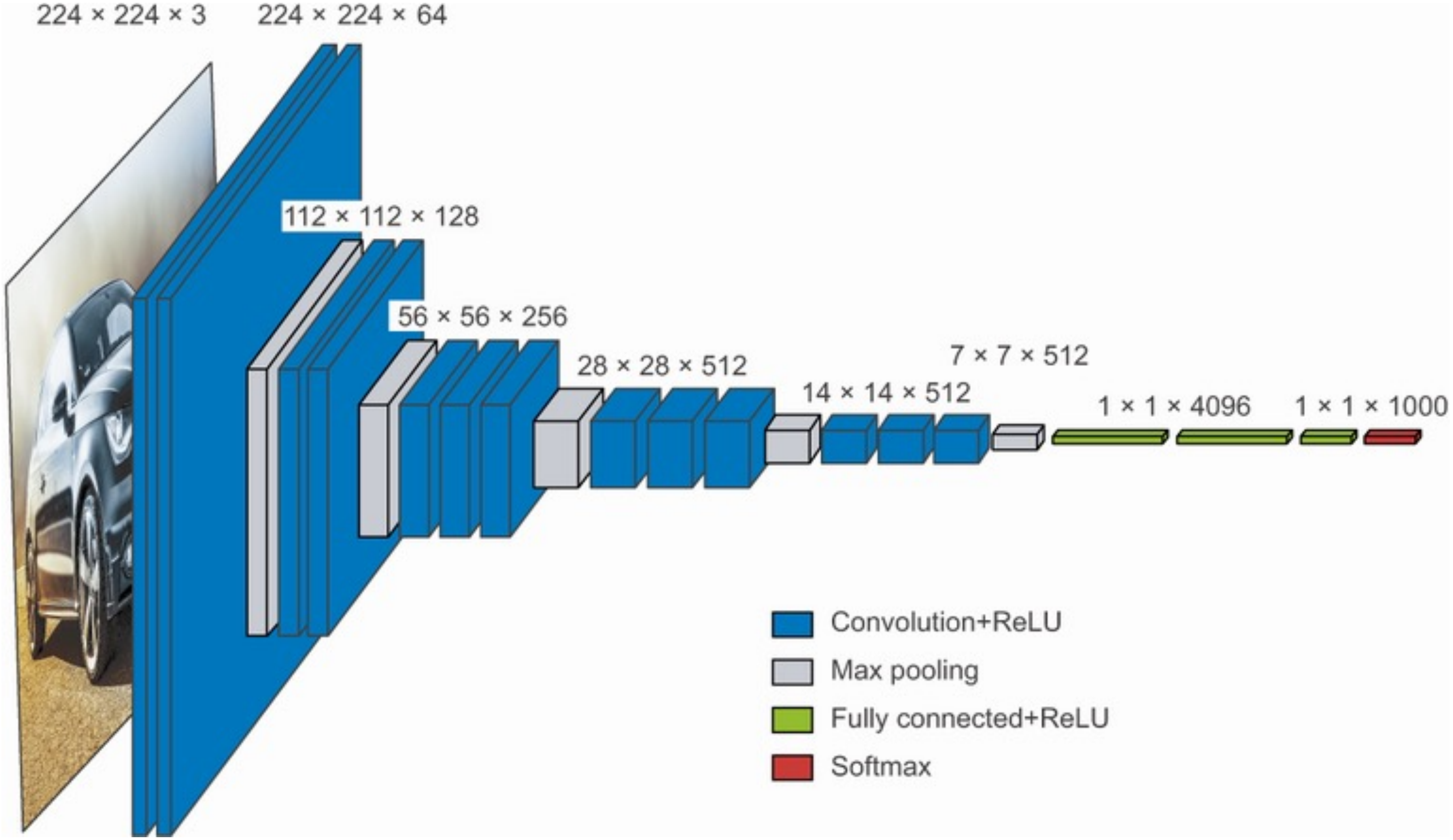
Image-to-Image Prediction Tasks

- Image segmentation, Image super-resolution.
- Transpose convolution operations for up-sampling.

Visualizing What a CNN is Learning

- Try to visualize what components of an image your model’s feature extractions are detecting.

Image-Classification Architectures



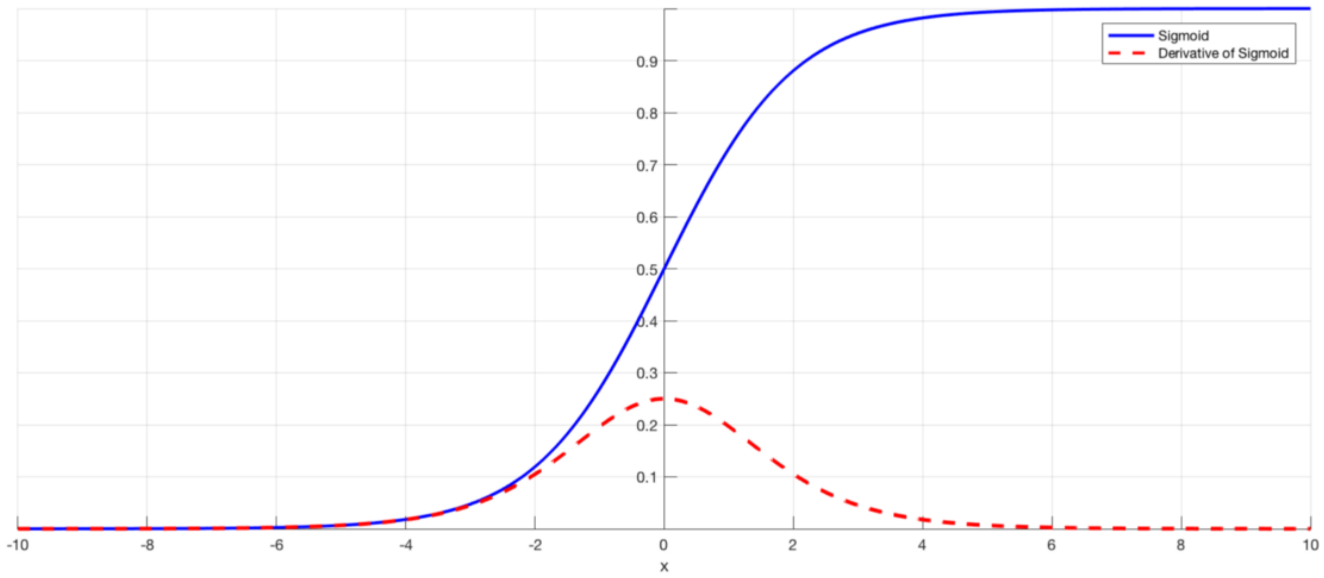
Vanishing Gradients

Recall That Certain Activations Make Vanishing Gradients More Likely

- As we approach 0 or 1 at a node's output, changes to input parameters have little impact…

However, Regardless of Activation This Can Be an Issue

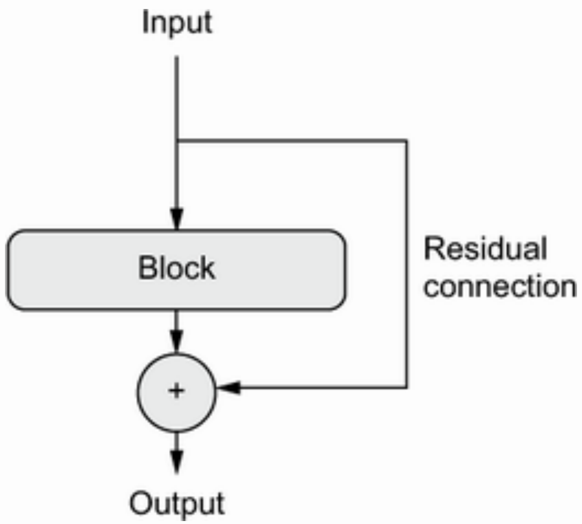
- For deep networks, parameters at front of network tend to have a much smaller influence on ultimate loss function. Accordingly, gradients for those early parameters can be very small.



Solution: Residual Connections

Provide a ’Short-cut’ From Loss Function to Front-end Weights

- We include feed-forward layers as usual, but we also add short-cut connections *around* the layers.
- We typically incorporate these residual connections either via an ‘Add’ layer or a ‘Concatenate’ layer.
- Note that conformity of the tensor shapes will be very important here – you’ll start encountering shape conformity errors if you are not careful!



Listing 9.2 Residual block where the number of filters changes

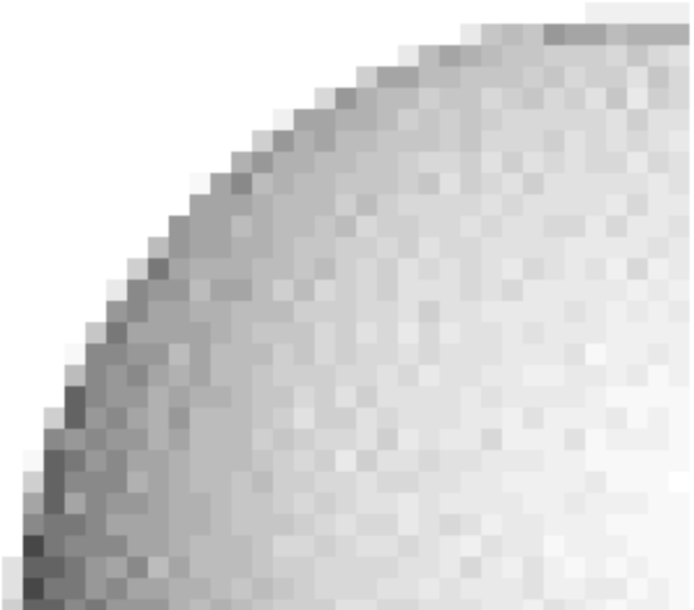
```
from tensorflow import keras
from tensorflow.keras import layers

inputs = keras.Input(shape=(32, 32, 3))
x = layers.Conv2D(32, 3, activation="relu")(inputs)
residual = x
x = layers.Conv2D(64, 3, activation="relu", padding="same")(x)
residual = layers.Conv2D(64, 1)(residual)
x = layers.add([x, residual])
```

Image-to-Image Prediction

Super-Resolution

- Take a high-resolution image, pixelate it, then try to predict the high resolution from the pixelated version.
- Q: What kind of activation and loss function will make sense in this task?



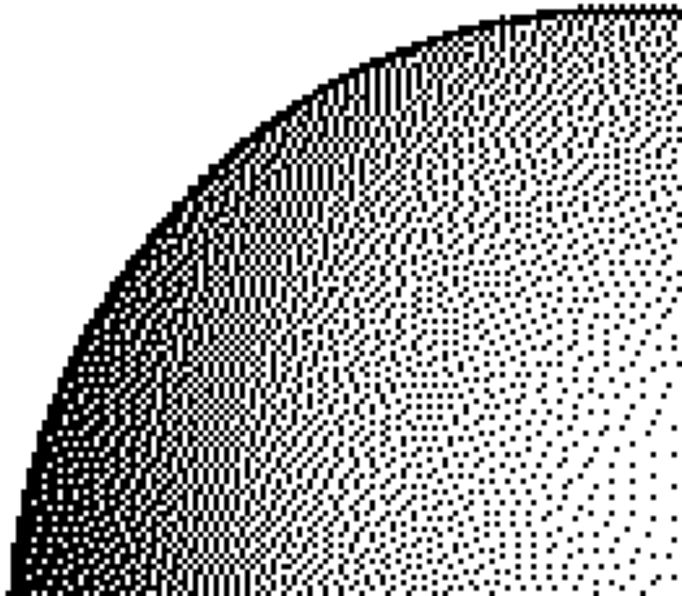
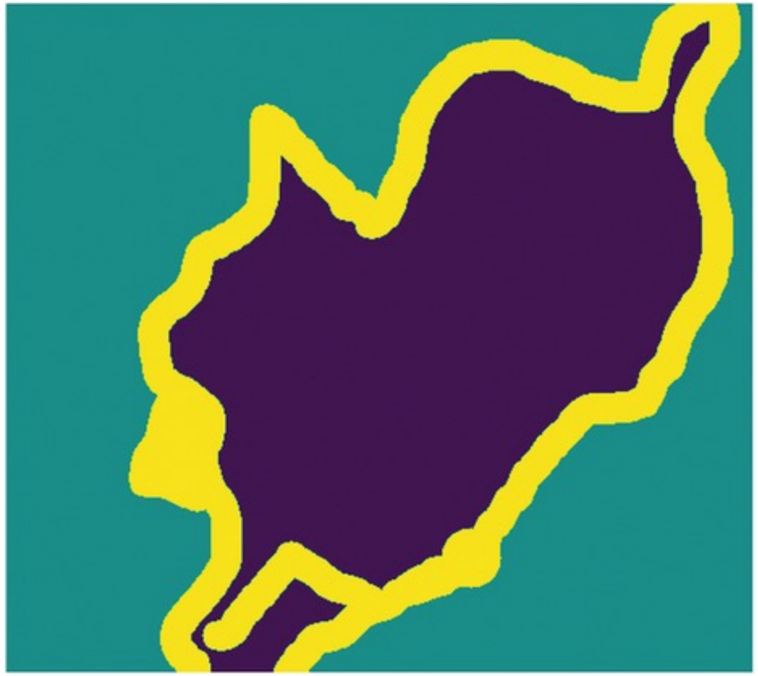


Image-to-Image Prediction

Image-Segmentation

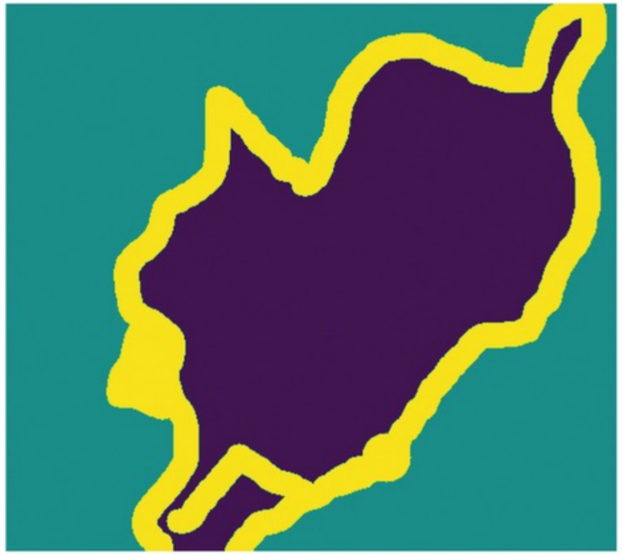
- Take an image and its segment mask, then try to predict the segment associated with each pixel from the original picture.
- Q: What kind of activation function and loss function will make sense in this task?



Topology for Image-to-Image

Auto-Encoder Architecture

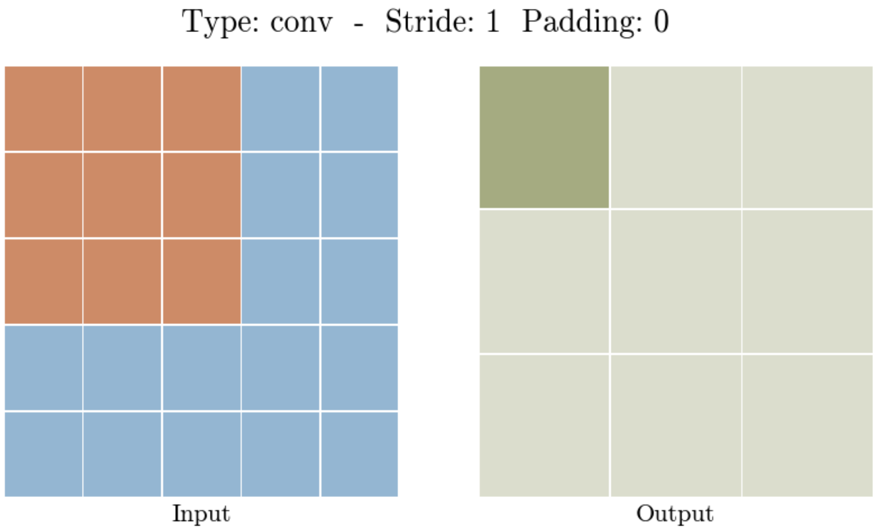
- Down-sample and then Up-sample back to same dimensionality
- We do not use down-sample using pooling (because they force attention to the whole image as we learn higher level features). Instead, we use larger strides. This enables 'dimensionality' reduction while maintaining a focus on local portions of the image.
- We then 'up-sample' back to the original dimensionality using a transpose of the convolutional operation. This is a form of autoencoder architecture.

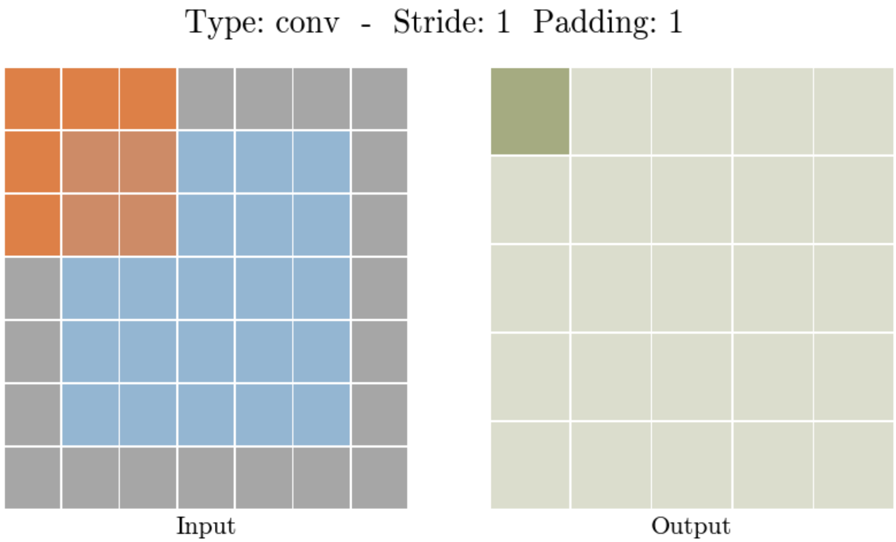


Convolution

Recall What A Convolution Is…

- It’s a down-sampling approach (it compresses information)

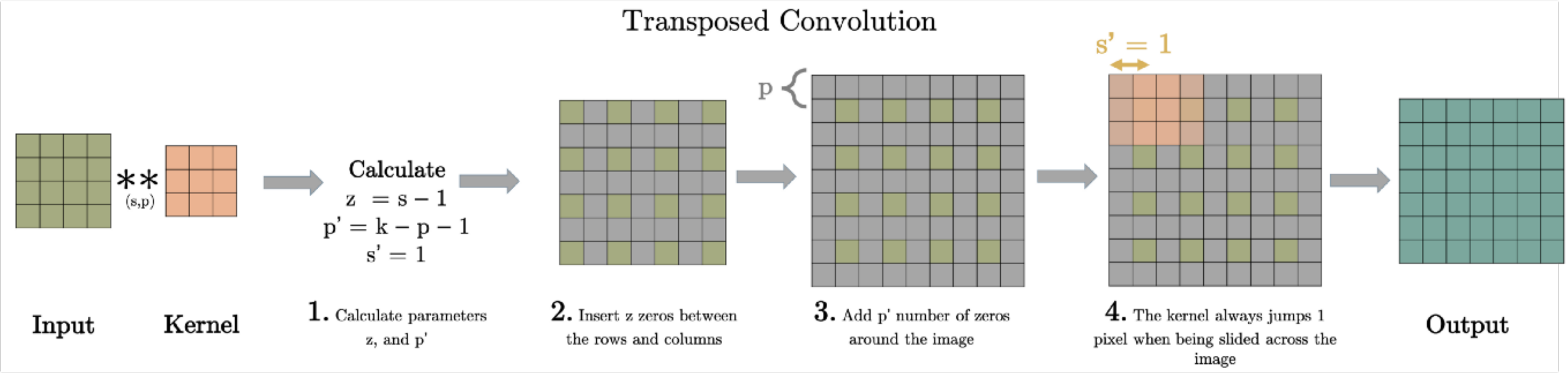




Inverse Convolution

Inverse of a Convolution

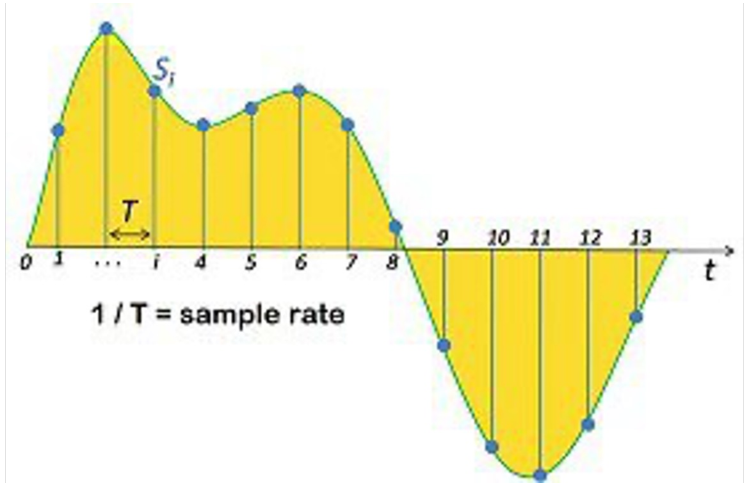
- Instead of Input (+padding) * filter (+stride) = output, this calculates the inverse operation, to up-sample.
- Here, s is the Convolution stride, p is the Convolution padding, k = is the Convolution kernel width/height; thus, s' is the stride of the transpose convolution (always 1), z = padding around each pixel element, p' is the final padding around the image for the transpose operation.



CNN for Audio

Same Sequence Concepts Work for Audio Data

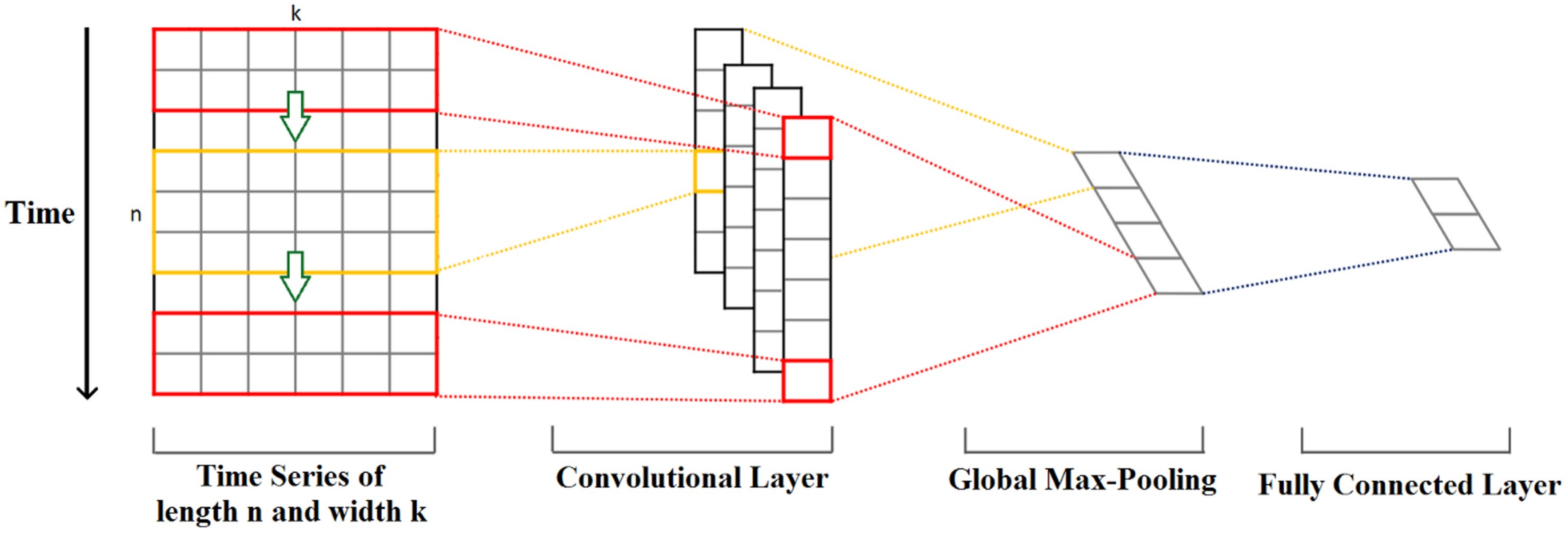
- Audio files are just sequences of numeric values (amplitude), possibly two if it was recorded in stereo.
- Once we recognize this, we realize we can predict things about audio sequences too!



(Temporal) 1D Convolution

1D Convolution Accomplishes Same Goal as 2D

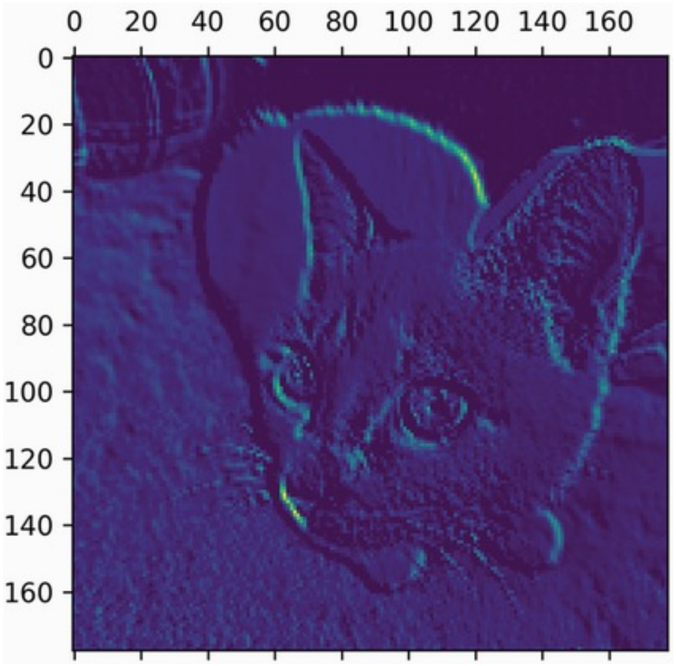
- It only considers arrangement of features in one dimension (temporal ordering).
- Compresses into shorter sequences, across the entire set of features (just as 2D Conv compresses matrices into smaller matrices, across the entire set of input channels (e.g., RGB).



Visualizing What a CNN is Learning

Visualizing Filter Activations in a Layer

- For a given layer, we plot the 2D feature map for each channel (filter). Each one will capture independent features of the image – plotting each layer's activation (output) can show us what pixels are being identified in the pictures to produce a prediction.



Visualizing What a CNN is Learning

Class Activation Maps

- Generate a heatmap on the original input image that indicates what ‘components’ (feature maps) of the input image most drive the final classification assignment based on ‘steepest’ gradients
- The idea here is to calculate the gradient of the output class w.r.t. mapped features from the input image and then highlight the pixel where the high-gradient features were most present.

Questions?