Beyond CNNs: Encoded Context for Image Inpainting with LSTMs and Pixel CNNs

Taneem Ullah Jan   Ayesha Noor

Image Inpainting?

Reconstructing the missing or damaged parts within the image

Some Common and Low Level Methods

- Convolutional Neural Networks (CNN in Raw form)
- Pixel-CNN
- Generative Adversarial Network (GAN and or forms)(High Level)
- Other methods based on Autoencoders and Attention Mechanism (High Level)

Pixel-CNN

- In CNN, the images are not considered in a sequence, instead we take the whole image as an input
- Then we apply filters using Conv Layers, Max/Avg Pooling, and Activation Function (if any in FC layers)
- In Pixel-CNN, we're doing the same but using the Masked Version
- That means we're masking certain values and apply filters to others
- Then those values are fed into fully connected layers to predict probabilities (results)

Our Model Extends Pixel-CNN :: Row Flattened LSTM

- Flattened the rows and append to each other to make a 1D array and then fed into an LSTM Network Architecture
- In this extended version of Pixel-CNN, the probability is computed based on each pixel from preceding rows
- That's why we called it, Row Flattened LSTM

Row Flattened LSTM
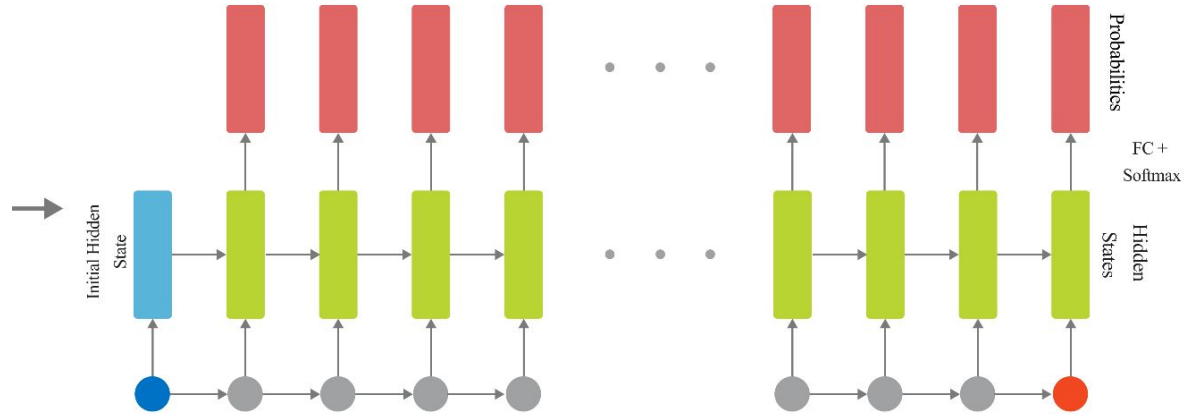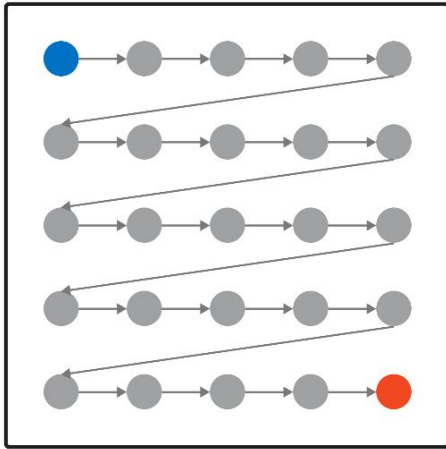
Result

On the left are Ground Truths, while on the right are the reconstructed image

Result Comparison :: L2 Loss

Model	L2 Loss
Row Flattened LSTM	**7.63**
Pixel CNN	6.98
Wasserstein GAN	4.26
CNN	7.49

Thanks