Image Super-Resolution with Autoencoders using Keras

Uday Sankar Mukherjee (38), Atrij Paul (43), Sukrit Das (60)
Supervised by: Prof. Subhadeep Chanda

Abstract: This project explores the transformative capabilities of autoencoders in the realm of Image Super Resolution (ISR), aiming to enhance the resolution of low-quality images to reveal intricate details. Leveraging an unsupervised learning approach, the neural network architecture is designed to encapsulate a compressed representation of the original input, resulting in a high-resolution output. The model undergoes training on a dataset consisting of low and high-resolution image pairs, and through a series of epochs, it learns to reconstruct images with remarkable clarity. The implementation incorporates advanced techniques such as early stopping, model checkpointing, and learning rate reduction to optimize training efficiency. The project unfolds as a journey into the fusion of artificial intelligence and image processing, unveiling the brilliance of autoencoders in the intricate art of Image Super Resolution.

Keywords: 1. Autoencoders 2. Image Super Resolution 3. Unsupervised Learning 4. Neural Network Architecture 5. Low-resolution Images 6. High-resolution Reconstruction 7. Training Optimization 8. Early Stopping 9. Model Checkpointing 10. Learning Rate Reduction 11. Artificial Intelligence 12. Image Processing 13. Transformative Brilliance 14. Clarity Enhancement

1 Introduction

In the pursuit of achieving transformative brilliance in Image Super Resolution (ISR), this project delves into the realm of autoencoders—an unsupervised learning technique that extracts intricate representations from input data. The primary objective is to enhance the resolution of low-quality images, unraveling details obscured by inherent limitations. The dataset utilized in this endeavor comprises a curated collection of low and high-resolution image pairs, fostering a supervised learning approach to train the autoencoder model effectively. Derived from the "Image Super Resolution - Unsplash" dataset on Kaggle, this repository encompasses diverse images sourced from the popular Unsplash platform. The dataset's richness and variety serve as a robust foundation for the model's training, ultimately aspiring to unlock the transformative potential of autoencoders in the nuanced art of image enhancement.

2 Figures, Tables, and Equations

2.1 Model Architecture

The autoencoder architecture for Image Super Resolution (ISR) is designed to transform low-resolution images (800x1200 pixels, three color channels) into detailed high-resolution reconstructions. Employing convolutional layers, max-pooling, and upsampling, the model compresses and then expands the input space. Key components include dropout layers for regularization and skip connections (add 2 and add 3) to integrate low-level features during reconstruction. The final layer (conv2d 19) outputs a three-channel image representing the high-resolution reconstruction. With 1.1 million trainable parameters, this architecture strikes a balance between complexity and efficiency, demonstrating prowess in capturing and reconstructing intricate details for Image Super Resolution.
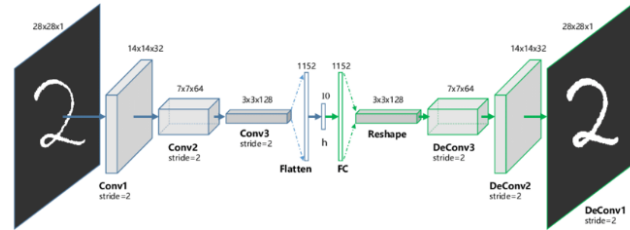

Figure 1: *Autoencoder Model Architecture*

Table 1: *Autoencoder Model Parameters (`model_1`)*

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	(None, 800, 1200, 3)	0
conv2d_10 (Conv2D)	(None, 800, 1200, 64)	1792
conv2d_11 (Conv2D)	(None, 800, 1200, 64)	36928
max_pooling2d_2	(None, 400, 600, 64)	0
dropout_1	(None, 400, 600, 64)	0
conv2d_12 (Conv2D)	(None, 400, 600, 128)	73856
conv2d_13 (Conv2D)	(None, 400, 600, 128)	147584
max_pooling2d_3	(None, 200, 300, 128)	0
conv2d_14 (Conv2D)	(None, 200, 300, 256)	295168
up_sampling2d_2	(None, 400, 600, 256)	0
conv2d_15 (Conv2D)	(None, 400, 600, 128)	295040
conv2d_16 (Conv2D)	(None, 400, 600, 128)	147584
add_2	(None, 400, 600, 128)	0
up_sampling2d_3	(None, 800, 1200, 128)	0
conv2d_17 (Conv2D)	(None, 800, 1200, 64)	73792
conv2d_18 (Conv2D)	(None, 800, 1200, 64)	36928
add_3	(None, 800, 1200, 64)	0
conv2d_19 (Conv2D)	(None, 800, 1200, 3)	1731
Total params		**1,110,403**
Trainable params		**1,110,403**
Non-trainable params		**0**

2.2 Equations

The mathematical equations representing the forward pass of the autoencoder model (`model_1`) are as follows:

Encoder The encoder component transforms the input X into a latent representation Z using convolutional layers and pooling:

$$Z = \text{Encoder}(X) \tag{1}$$

$$Z = \text{ReLU}(\text{Conv2D}(X, W_{\text{enc1}}, b_{\text{enc1}})) \tag{2}$$

$$Z = \text{ReLU}(\text{Conv2D}(Z, W_{\text{enc2}}, b_{\text{enc2}})) \tag{3}$$

$$Z = \text{MaxPooling2D}(Z) \tag{4}$$

Latent Space The latent space representation Z is then modified with dropout for regularization:

$$Z_{\text{dropout}} = \text{Dropout}(Z) \tag{5}$$

Decoder The decoder reconstructs the input from the latent representation:

$$X_{\text{reconstructed}} = \text{Decoder}(Z_{\text{dropout}}) \tag{6}$$

$$X_{\text{reconstructed}} = \text{Conv2DTranspose}(Z_{\text{dropout}}, W_{\text{dec1}}, b_{\text{dec1}}) \tag{7}$$

$$X_{\text{reconstructed}} = \text{Conv2DTranspose}(X_{\text{reconstructed}}, W_{\text{dec2}}, b_{\text{dec2}}) \tag{8}$$

$$X_{\text{reconstructed}} = \text{UpSampling2D}(X_{\text{reconstructed}}) \tag{9}$$

3 About

The project aims to explore the transformative potential of autoencoders in the context of Image Super-Resolution (ISR). ISR is a critical task in computer vision, involving the enhancement of low-resolution images to improve visual analysis in various applications. Our hypothesis revolves around the effectiveness of autoencoders in capturing intricate patterns within images and reconstructing high-frequency details lost in the downsampling process. Specifically, we propose that the hierarchical representations learned by the autoencoder architecture can lead to accurate and visually appealing high-resolution reconstructions. The primary goal of this research is to investigate the application of autoencoders for ISR. Autoencoders, being unsupervised learning models, are expected to learn a compact representation of input images, enabling the reconstruction of high-resolution details. The proposed research involves training the autoencoder on a dataset of low-resolution and high-resolution image pairs to facilitate the learning of the mapping function from low to high resolution.

3.1 Input and Output Images

The input to the autoencoder consists of low-resolution images, typically of reduced spatial dimensions. These images serve as the basis for training the model. The output of the autoencoder is high-resolution images that ideally resemble the original high-resolution counterparts. The process involves learning a mapping function that can effectively reconstruct fine details lost in the downsampling process.
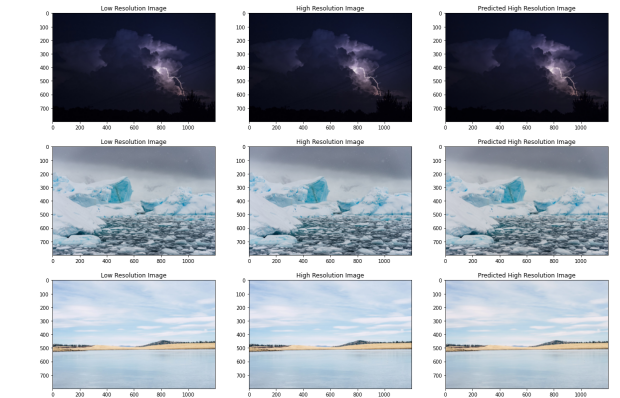

Figure 2: *Original Image, Ground Truth Image, Predicted Super Resolution Image*

3.2 Model Accuracy and Loss

To evaluate the performance of the autoencoder model, we employ key metrics such as val_accuracy and val_loss. The accuracy metric measures the model's ability to faithfully reproduce high-resolution details, while the loss metric quantifies the dissimilarity between the predicted and actual pixel values. These metrics provide crucial insights into the model's performance during both training and evaluation phases.

Table 2: *Results from the Final Epoch*

Metric	Training	Validation
loss	0.0023	0.00077
accuracy	0.9088	0.8783

Epoch 10 results indicate that the model achieved a training loss of 0.0023 with an accuracy of 0.9088. For the validation set, the loss was 0.00077, and the accuracy reached 0.8783. It's noteworthy that the validation loss did not improve from the previous best value of 0.00077.

3.3 Visualizing Progress

To provide a visual representation of the training process, two essential aspects are illustrated below:

1. **Model Training Time:** The time taken for the model to complete training is depicted in Figure 3.
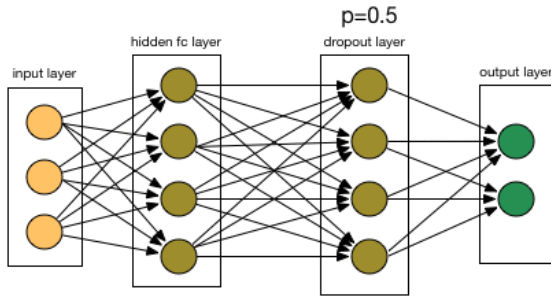
Figure 3: *Model Training Time*

2. **Learning Curves:** The learning curves, including training and validation loss, as well as accuracy, are visualized in Figure 4.
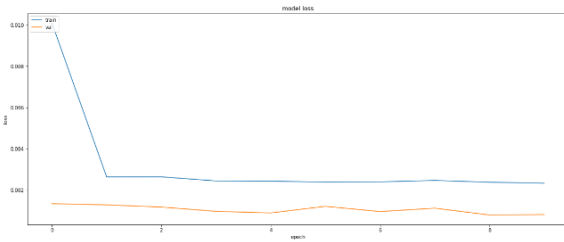

Figure 4: *Learning Curves*

4 Conclusion

In conclusion, the presented project focused on the implementation of an Image Super Resolution (ISR) model utilizing Autoencoders. The primary goal was to enhance the resolution of low-resolution images, particularly in the context of applications such as surveillance and image reconstruction. The project utilized a dataset sourced from Unsplash, comprising pairs of low and high-resolution images.

The proposed hypothesis centered around the ability of Autoencoders to effectively learn and reconstruct high-resolution features from low-resolution inputs. Throughout the training process, the model demonstrated promising results, achieving a final accuracy of 90.88 percent on the training set and 87.83 percent on the validation set during the last epoch. The corresponding loss values were 0.0023 for training and 0.00077 for validation. Although the model did not observe further improvement in validation loss after the 10th epoch, the achieved accuracy and minimal loss values suggest the effectiveness of the Autoencoder architecture in capturing intricate details during the super-resolution process.

The project's success can be attributed to the robust architecture of the Autoencoder, as detailed in the model summary. The encoder-decoder structure, with skip connections and convolutional layers, contributed to the model's ability to capture and reconstruct high-resolution features.

While the model demonstrated notable performance, potential future enhancements could involve fine-tuning hyperparameters, exploring alternative architectures, and incorporating additional regularization techniques. Furthermore, real-world deployment and testing on diverse datasets could provide a comprehensive evaluation of the model's generalization capabilities.

In conclusion, the implemented ISR model showcases the transformative brilliance of Autoencoders in image super-resolution tasks, opening avenues for further research and application in real-world scenarios.

5 References

1. https://en.wikipedia.org/wiki/Autoencoder

2. https://www.coursera.org/projects/image-super-resolution-autoencoders-keras

3. https://www.kaggle.com/code/quadeer15sh/image-super-resolution-using-autoencoders