Compression and Restoration of Images Sent Over Social Media

Arushi Singh
Shayla Luong
Aditi Agrawal

ABSTRACT

Many social media services such as Facebook, Whatsapp, Twitter allow users to upload and share an unlimited number of images through their online platform. In order to reduce the data size for storing, handling, and transmitting these images, the services perform a compression algorithm on the images when users first upload them for sharing. The compression algorithms produce smaller file sizes but at the cost of decreasing the image quality and introducing undesired degradations. This project experiment consists of two parts, one is in compressing the image using an autoencoder network and the other is in restoring the image using a high-performance image restoration model. The image compression step uses the Convolutional AutoEncoder (CAE) network and the restoration step uses the SwinIR model as the baseline. We successfully trained a CAE model and completed transfer-learning on a pre-trained SwinIR model. The results of CAE compression are sent to the new restoration model for decompression.

Keywords: Convolutional Autoencoder, image compression, image restoration, SwinIR

I. INTRODUCTION

Exchanging photos/videos over social media over the internet is very common these days. Most of us use some or the other form to send media over the internet. Unfortunately, we feel frustrated when the HD photos captured using good cameras sent over social media, are not what we captured, basically 'pixelated'?

When we send photos over the internet using messenger apps, the image quality decreases to a greater extent almost to 70% [1]. Most of these messenger apps use a compression technique called lossy compression to downsize the image in order to send more data over the internet. Lossy compression reduces a file by permanently eliminating certain information, especially redundant information. When the file is uncompressed, only a part of the original information is still there which causes the image quality to degrade.

This problem is tackled at two fronts. Once on the sender side. When the image is being sent over the internet, a "Deep Convolutional AutoEncoder-based Lossy Image Compression " technique is used for compression which has a good compression rate and retains important features of the image. Secondly on the receiver side, when the image is decompressed, the quality of the reconstructed image is better than other compression techniques. The quality of these reconstructed images is further enhanced using the super-resolution(SR) technique from “SwinIR: Image Restoration Using Swin Transformer".

II. LITERATURE SURVEY

One of the classical Computer Vision approaches employs the frequency domain to perform the image super-resolution task. The pioneering work from Tsai and Huang [2] aims at recovering high-resolution images from multiple shifted low-resolution images by using shift and aliasing properties of Fourier Transform. This first formulation assumes the input images contain no other degradations like noise and spatial blur during the down sampling process. Other early works [3,4] introduce regularization and other extensions to improve this method’s capability in handling more complicated degradations. These methods are based on estimating motions created by the physical camera mechanism that produces multiple frames for each image capture. Many future works with more success in image restoration are done in the spatial domain [5]. One early approach in this domain was through using interpolation and restoration. The procedure consists of three key parts. The low-resolution image frames are first aligned and placed onto a high-resolution image grid. Then, nonuniform interpolation formulas (e.g. Bicubic and Bilinear interpolation, weighted nearest neighbors [6]) are used to fill in missing pixel values. The final step was focused on classical deblurring and noise removal algorithms such as Wiener filtering and Gaussian noise distribution. Another classical approach uses different statistical algorithms such as maximum likelihood, Bayesian, and Gaussian Markov Random Field to generate high-resolution images [5]. Most of these classical approaches require knowledge of image priors for regularization and are limited in the amount of undesired image degradation that can be processed.

L Theis, W Shi, A Cunningham and F Huszar in[7] introduced a simple but effective way of dealing with non-differentiability in training autoencoders for lossy compression. Together with an incremental training strategy, this enabled them to achieve better performance than JPEG 2000 in terms of SSIM and MOS scores. Notably, this performance was achieved using an efficient convolutional architecture, combined with simple rounding based quantization and a simple entropy coding scheme. Here they aim at directly optimizing the rate-distortion trade off produced by an autoencoder.They propose a simple but effective approach for dealing with the non-differentiability of rounding-based quantization, and for approximating the non-differentiable cost of coding the generated coefficients.

[8] based on generative adversarial networks (GAN), which is a generator and discriminator network. In this method the generator is given an input of a low resolution image and it tries to generate a high resolution image from it. The discriminator takes the generated image from the generator as an input and computes loss by comparing it

to the original high resolution image. This loss is then further used to further improve the performance of the generator. These two networks compete with each other to generate an image which looks closer to the original high resolution image.

III. METHODOLOGY

For the enhancement of lossy compression at the sender's side, we have used Deep Convolutional AutoEncoder [9], as this is an improvement on state-of-the-art compression techniques. Once the compression is done, we work on restoring the image on the receiver's end. For the image restoration task, we are using SwinIR [10] model to upscale the image. The model has the capability of handling large-size images and able to model long-range dependency.

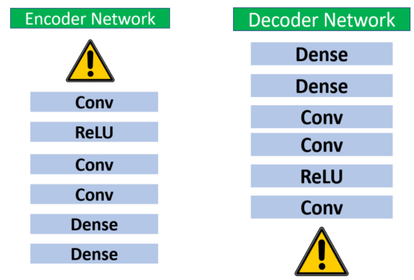
Deep convolutional autoencoder-based lossy image compression[8]: Z Cheng et al., in [8] mentions a new architecture which utilizes the advantages of convolutional autoencoder (CAE) to achieve a high coding efficiency. First, they design a novel CAE architecture to replace the conventional transforms and train this CAE using a rate-distortion loss function. Experimental results demonstrate that their method outperforms traditional image coding algorithms, by achieving a 13.7% BD-rate decrement on the Kodak database images compared to JPEG2000.

SwinIR: Image Restoration Using Swin Transformer [10]: In [10], the proposed SwinIR model consists of shallow feature extraction, deep feature extraction, and high-quality image reconstruction modules. The first feature extraction module uses a convolutional layer that is especially useful and simple for mapping the input image space to a higher dimensional space. The deep feature extraction step is composed of Swin Transformer layers, convolution layer, and residual connection. In the image reconstruction module, output features from shallow and deep extraction are aggregated and a sub-pixel convolution layer is used to up-sample the features. This proposed model achieves great results in the areas of classic image super-resolution, color image denoising, and JPEG compression artifact reduction. Transformer-based designs use a local self-attention mechanism that gives the capability of handling large-size images. It's also able to model long-range dependency by employing the shifted window scheme.

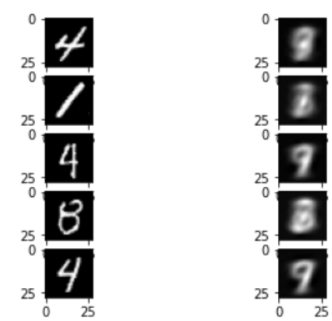
IV. IMPLEMENTATION DETAILS

Compression

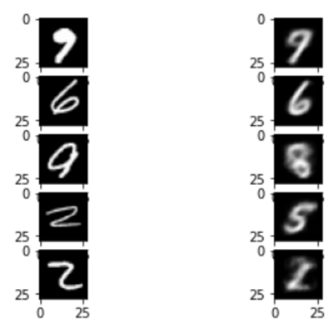
As part of implementing the Convolutional Autoencoder for compressing the image, we started with implementing a plain Autoencoder with two dense layers in encoder and similarly two dense layers in decoder. We used the MNIST data set to compress the image and decompress it. Activation function used is LeakyRelu. The autoencoder does a decent job in recreating the image, but in order to enhance it further, we implemented Convolutional Autoencoder(CAE). CAE consists of three convolutional layers and two dense layers with ReLU as the activation function. The Architecture of the CAE is mentioned below:



CAE performed much better than autoencoder in recreating the compressed images with better quality, though it took more time to train the model.The autoencoder took about 10-15 mins to train the model for 10 epochs where as the CAE took about 2 hours to train the model for 10 epochs. The results are as follows:

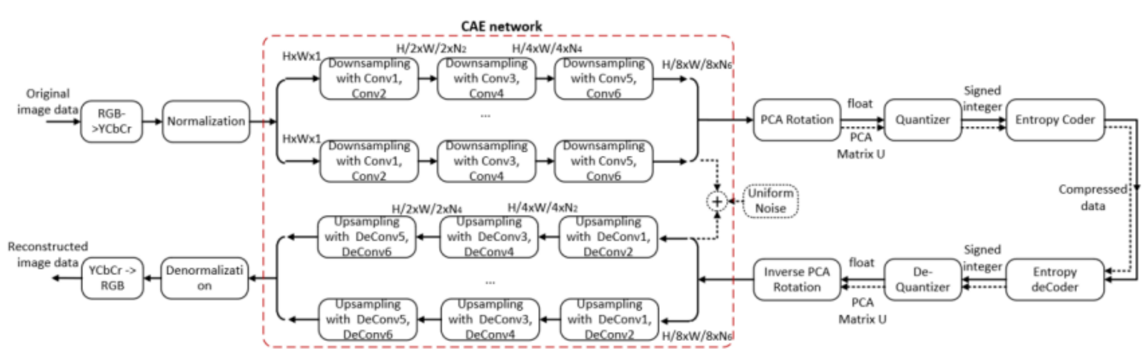


Decompressed image from the Autoencoder with two Dense Layer



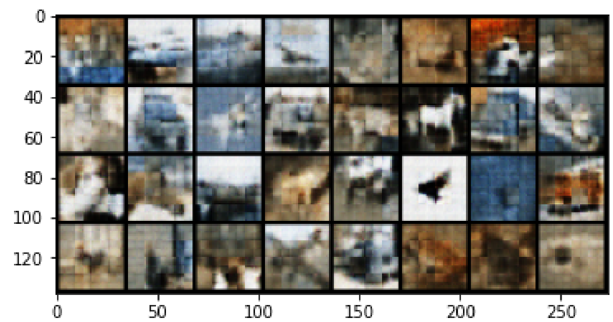
Decompressed image from the Convolutional Autoencoder.

The above implementations are for grayscale images, and now we tried the similar models using CIFAR10 dataset. As part of our implementation we used pytorch this time and used three convolutional layers in encoder and three convolution layers in decoder. The activation function used is relu. We removed the dense layers for the coloured images. The optimizer used is Adam and the learning rate is 0.0001. The architecture for the model is as follows.



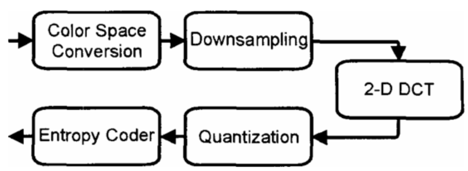
We implemented the RGB to YCrCb conversion but it resulted in a change of color after decompressing the image. As a part of the future work, we will inspect it further. For now we are feeding RGB images to the model. The execution time for running the model for 5 epochs is 7 mins. The result for the same are as follows:

Original 32 images of size 32x32x3



Reconstructed Images by using CAE

In order to depict the final results and compare it with JPEG compression , we have implemented JPEG compression as well. Where the architecture of the JPEG compression is as follows:

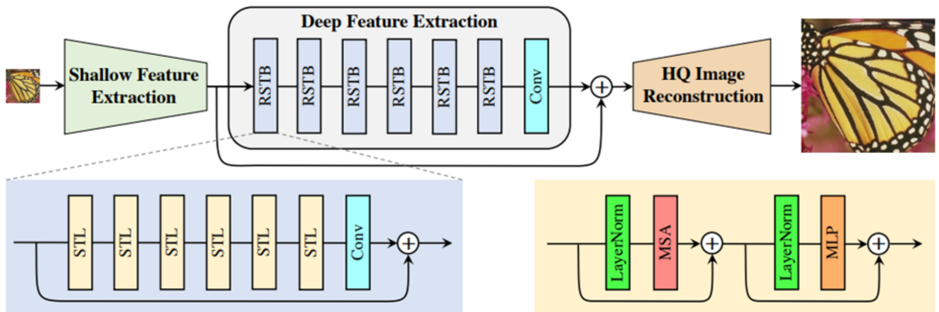


We compressed the image using JPEG compression and reconstructed it and sent the reconstructed image to the SWIN IR model to further enhance its resolution. As part of our comparison we sent five images of size 32x32x3 into the trained CAE model and sent the reconstructed image to the SWIN IR model. The results of the comparison as mentioned in the results section.

Restoration

SwinIR works on different aspects of image restoration, namely, classical/lightweight image SR, real-world image SR, color/grayscale image denoising and JPEG compression artifact reduction. For our implementation we have used the pre-trained mode of real-world image SR. We have retrained the model on our data. Which is 50 high resolution images and 10 validation images. The real-world image super resolution uses high resolution images and generates low quality images from the high quality image using BSRGAN. BSRGAN adds blur twice, down samples twice, adds gaussian noise, jpeg noise and camera sensor noise. The model is trained on these degraded images and the original high resolution image is used as ground truth. We have used a combination of L1 loss and perceptual loss for training the mode.

Further, Adam optimizer has been used as optimizer. We ran training and validation for 5 epochs (due to computation restriction). SwinIR architecture is below.



(a) Residual Swin Transformer Block (RSTB)
(b) Swin Transformer Layer (STL)

V. RESULTS AND DISCUSSION

Compression

The metrics used to evaluate the compression and restoration of the image are SSE, PSNR, SSIM. The dataset used for this evaluation is CIFAR10 images which are of size 32x32x3. As part of evaluating how much better the image gets reconstructed after compression we are checking the PSNR values and SSIM values for the original image vs JPEG compression and original image vs Convolutional Autoencoder compression. Bigger the PSNR value the better is the compression.
The SSIM values range between -1 to 1. Closer the value to 1,better is the structural similarity between the images.

Images	JPEG_PSNR	CAE_PSNR
img1	25.133571730052722	68.80298894094759
img2	25.133571730052722	72.0766678288864
img3	23.492463587258207	71.06848423751096
img4	23.782298740309713	69.24729347565386
img5	24.498634719616142	70.9987425569537

Images	JPEG_SSIM	CAE_SSIM
img1	0.7324647203460863	0.9985165564310678
img2	0.8404630172197501	0.9985165564310678
img3	0.9898393374143969	0.9898393374143969
img4	0.6912463978940243	0.9980587167189803
img5	0.7678400774761748	0.9957950897580261

As we can see, the PSNR values of the decompressed image from the CAE model are more than 50% higher than the PSNR values of the images from JPEG compression. SSIM values on the other hand are closer to 1 for the decompressed images from the CAE model but decompressed images from JPEG perform equally well.
These two set of images are now sent to SWIN IR model to improve the resolution on the sender's end.

Restoration

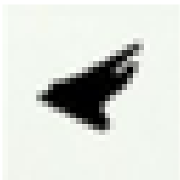
We ran training on 50 high resolution images for train and 10 images for validation. We ran 5 epochs and got the following PSNR after each epoch.

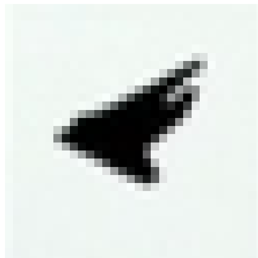
Epoc	AVG PSNR
1	24.309989679055892
2	26.74641404859804
3	26.69878608011634
4	27.317388404477526
5	26.18571235406673

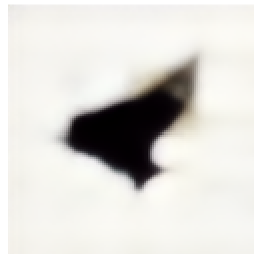
Following is the result of applying super resolution using our model on a degraded image(Might have to zoom to see the difference).

Final Result

Following is an example of an original photo of size 32x32x3 , followed by JPEG compression and decompression of the image, followed by compression of the image using CAE and restoration using SWIN IR. The restored image from our model is of size 128x128x3.

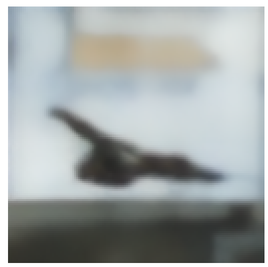

Original Image


JPEG Compression


CAE Compression + SWIN IR Restoration


Original Image


JPEG Compression


CAE Compression + SWIN IR Restoration

REFERENCES

[1] V. Punj, “Whatsapp update on ios let’s beta users send HD photos and how it works.” https://www.livemint.com/technology/apps/whatsapp-update-on-ios-lets-beta-users-send-hd-photos-how-it-works-11627203919427.html, 2021.

[2] R. Y. Tsai and T. S. Huang. “Multipleframe image restoration and registration.” In *Advances in Computer Vision and Image Processing*, pages 317–339. Greenwich, CT: JAI Press Inc., 1984.

[3] Sean Borman and Robert L. Stevenson. “Super-resolution from image sequences - A review.” In *Proceedings of the 1998 Midwest Symposium on Circuits and Systems*, pages 374–378, 1998.

[4] Z. Lin and H.-Y. Shum. “Fundamental limits on reconstruction-based superresolution algorithms under local translation.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):83–97, 2004.

[5] J. Yang and T. Huang. “Image super-resolution: Historical overview and future challenges.” 10.1201/9781439819319-1, 2017.

[6] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda. “Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames.” *IEEE Transactions on Instrumentation and Measurement*, 49(5):915–923, 2000.

[7] T. Lucas, W. Shi, A. Cunningham, and F Huszár. "Lossy image compression with compressive autoencoders." in *arXiv preprint arXiv:1703.00395*, Mar. 2017. Available: https://arxiv.org/pdf/1703.00395v1.pdf

[8] Wei, Zihao, et al. "A-ESRGAN: Training Real-World Blind Super-Resolution with Attention U-Net Discriminators." in *arXiv preprint arXiv:2112.10046*, 2021. Available: https://arxiv.org/abs/2112.10046

[9] Z. Cheng, H. Sun, M.Takeuchi and J. Katto. “Deep convolutional autoencoder-based lossy image compression.” In *IEEE Picture Coding Symposium (PCS)*, pages 253-257, 2018

[10] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte. “SwinIR: Image Restoration Using Swin Tranformer,” in *arXiv:2108.10257v1 (eess.IV)* 23 Aug 2021. [Online]. Available: https://arxiv.org/abs/2108.10257