Image Compression on Heterogeneous Images Using Convolutional Neural Networks

[1]Anix Jugal D'Cunha, [2]Sushant Mangasuli, [3]Vaibhavi Sahukar, [4]Afrah Saleem [5]Akshatha Sagar

Alva's Institute of Engineering and Technology,
Visvesvaraya Technological University, Mangalore, India

***Abstract*: Data compression has always been a key concern during transmission. Interests on image processing has been increased enormously from last decades. As a result different compression techniques has been introduced and purposed. Lossy image compression algorithms are used widely where some information gets lost during compression resulting high compression. However, we pay for their high compression rate with visual artifacts degrading the user experience. Deep convolutional neural networks (CNN) have become a widely used tool to address computer vision tasks very successfully.**

I. INTRODUCTION

This chapter gives a brief introduction about the Machine Learning and project entitled "Image Compression Using Convolutional Neural Networks". With the development of multimedia technology and communication technology, multimedia entertainment has played an important role in people's daily lives. Pictures and videos take up the main part of multimedia entertainment. It brings austere challenge to store and transmit those data, and puts forward higher requirement on the limited-bandwidth internet, especially for large and high-quality digital images. The limited bandwidth of internet greatly restricts the development of image communication, and thus the image compression technology has been more and more aroused people's attention. The purpose of image compression is to represent and transmit the original large image with minimal bytes, and to restore the image with not-so-bad quality. Image compression reduced the burden of image storage and transmission on the network, and achieved rapid real-time processing online. The information of an image is fixed, but the different representations of the image lead to different changes in the amount of data stored in the image. So in the representation with larger amount of data, some data is useless or represent the information that is represented by other data, they are irrelevant or redundant. The main purpose of image compression is to compress the image by removing redundant or irrelevant information, and to store and transmit digital compressed data on a low bandwidth network. Image compression methods can be classified into two kinds: one may lose information during compression, and the other one can keep full information, that is, lossless coding methods and limited-distortion coding methods. Lossless coding methods will not suffer loss of information after compressing images, yet without a good compression ratio. The basic principle of this kind of methods is: an image consists of features, using the statistical features of the image, if a feature appears many times in the image, it will be encoded in shorter bits, and if a feature appears only once or limited times, it will be encoded in longer bits. And a complete image is always composed of a large number of repeated features. According to that, the image will be represented by many short-bits coding features and little long-bits coding features. On the basis of guaranteeing the image quality after compression, limited-distortion coding methods maximize the compression ratio. The original image and the compressed image looks very similar though some information has changed. The normal used limited-distortion coding methods are: the predictive coding method, the transform coding method and the statistical coding method. To replace the transform and inverse transform in traditional codecs, we design a symmetric CAE structure with multiple down sampling and up sampling units to generate feature maps with low dimensions. We optimize this CAE using an approximated rate-distortion loss function. The features of images can be learned automatically using deep learning models, rather than proposed manually. Suitable features can improve the performance of image recognition. Over the past years, features of images were always specified manually that depended on the designers' prior knowledge, and the number of features were very limited. Deep learning models can automatically learn unlimited number of features automatically. A good feature-extraction method is a prerequisite for optimization of image processing. In this study, we proposed a model to compress images. Based on AE, a multi-layer model is constructed. An image is put into the first layer and the output data from different level of layers reconstruct the original image in different level of comprehension. If the size of the output data from an arbitrary layer is less than the size of the original image, the representation in this layer is a compression representation. Because the model has more than one hidden layer whose neurons are less than the input layer's, the model can achieve multiple levels of features, and each level of features represent a compressed image. So, multiple compression ratio can be obtained using this model.

II. PROPOSED CONVOLUTIONAL AUTOENCODER BASED IMAGE COMPRESSION

The features of deep learning can be learned automatically using deep learning models, rather than proposed manually. Suitable features can improve the performance of image recognition. Over the past years, features of images were always specified manually that depended on the designers' prior knowledge, and the number of features were very limited. Deep learning models can automatically learn unlimited number of features automatically. A good feature-extraction method is a prerequisite for optimization of image processing. In this study, we proposed a model to compress images. Based on AE, a multi-layer model is constructed. An image is put into the first layer and the output data from different level of layers reconstruct the original image in different level of comprehension. If the size of the output data from an arbitrary layer is less than the size of the original image, the representation in this layer is a compression representation. Because the model has more than one hidden layer whose neurons are less than the input layer's, the model can achieve multiple levels of features, and each level of features represent a compressed image. So, multiple compression ratio can be obtained using this model. The rapid development of the Internet of Things (IoT) has greatly facilitated people's lives, and it has also led to an explosive increase in the amount of data transmitted by networks. The types of network services have developed from the original text and voice signals to image and

Video signals, which brings convenience for the transmission of information and also continuously improves the requirements for data transmission and storage. Therefore, in order to reduce the volume of images during transmission and storage to improve network transmission efficiency, obtaining a better recovery quality through a smaller compression size has long been the focus of research in the image field. The objective of image compression is to reduce the redundancy of the image and to store or transmit data in an efficient form. Image compression is to reduce the data volume of and to achieve a low bit rate in the digital representation of images without perceived loss of image quality.

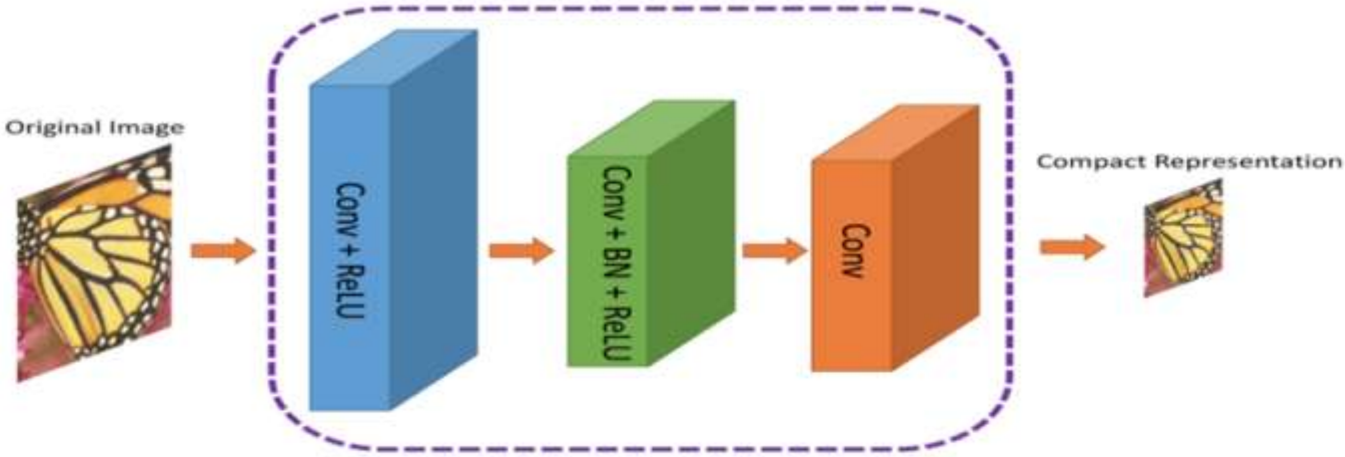

Fig. 1: Image Compression using CNN

A. Dataset collection

Different signs of input images are collected from web camera. There are 1000 images considered dataset for training. The images taken are colored images of size 256 * 256 image dimensions.

B. Image pre-processing

For each image in the provided train folder, patches are cut out of the image (patch size -> 50*50 with stride of 10). Each patch is randomly rotated and compressed into quality in range of [5,10,50,90,100]. The compressed patch is subtracted from the original patch to get the artifact data (residual data). The compressed patch becomes the input to the network and the artifact is learnt by the model (o/p). Once the model is trained a compressed image is given to the network and the artifact produced by the network is subtracted from the compressed image.

C. CNN Algorithm

DnCNN, and extend it for handling several general image denoising tasks. Generally, training a deep CNN model for a specific task generally involves two steps: (i) network architecture design and (ii) model learning from training data. For network architecture design, we modify the VGG network [26] to make it suitable for image denoising, and set the depth of the network based on the effective patch sizes used in state-of-the-art denoising methods. For model learning, we adopt the residual learning formulation, and incorporate it with batch normalization for fast training and improved.

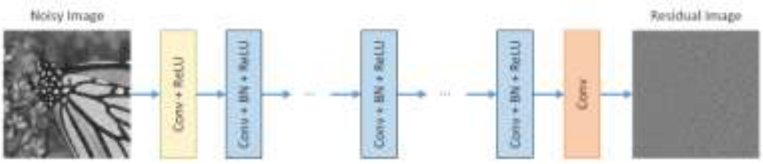

Fig 2: Deep Architecture

Deep Architecture: Given the DnCNN with depth D, there are three types of layers, shown in above figure with three different colors.

(i) Conv+ReLU: for the first layer, 64 filters of size 3X3Xc are used to generate 64 feature maps, and rectified linear units (ReLU, max(0,.)) are then utilized for nonlinearity. Here c represents the number of image channels, i.e., c = 1 for gray image and c = 3 for color image.

(ii) Conv+BN+ReLU: for layers 2, 64 filters of size 3X3X64 are used, and batch normalization is added between convolution and ReLU.

(iii) Conv: for the last layer, c filters of size 3X3X64 are used to reconstruct the output.

D. Compression Output

The image input is given and the trained model.h5 file is loaded to predict the output image. The output image is compressed image and we also get the residual image from the corresponding function. The restored image is generated through compressed image - residual image. The output images are stored in results folder.

E. literature survey

This chapter gives the overview of literature survey. This chapter represents some of the relevant work done by the researchers. Many existing techniques have been studied by the researchers on sign language prediction problem, few of them are discussed below.

I. An efficient block based lossless compression of medical images

Venugopal, Dhanasekar & Mohan, S. & Avaninathan, Sivanantharaja. (2015) International Journal for Light and Electron Optics. 127. 10.1016/j.ijleo.2015.10.154.

Medical images play a significant role in diagnosis of diseases and require a simple and efficient compression technique. This paper proposes a block based lossless image compression algorithm using Hadamard transform and Huffman encoding which is a simple algorithm with less complexity. Initially input image is decomposed by Integer wavelet transform (IWT) and LL sub band is transformed by lossless Hadamard transformation (LHT) to eliminate the correlation inside the block. Further DC prediction (DCP) is used to remove correlation between adjacent blocks. The non-LL sub bands are validated for Non-transformed block (NTB) based on threshold. The main significance of this method is it proposes simple DCP, effective NTB validation and truncation. Based on the result of NTB, encoding is done either directly or after transformation by LHT and truncated. Finally all coefficients are encoded using Huffman encoder to compress. From the simulation results, it is observed that the proposed algorithm yields better results in terms of compression ratio when compared with existing lossless compression algorithms such as JPEG 2000. Most importantly the algorithm is tested with standard non medical images and set of medical images and provides optimum values of compression ratio and is quite efficient.

II. Data-driven sparsity-based restoration of JPEG-compressed images in dual transform-pixel domain

X. Liu, X. Wu, J. Zhou and D. Zhao, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 5171-5178, doi: 10.1109/CVPR.2015.7299153.

Arguably the most common cause of image degradation is compression. This papers presents a novel approach to restoring JPEG-compressed images. The main innovation is in the approach of exploiting residual redundancies of JPEG code streams and sparsity properties of latent images. The restoration is a sparse coding process carried out jointly in the DCT and. pixel domains. The prowess of the proposed approach is directly restoring DCT coefficients of the latent image to prevent the spreading of quantization errors into the pixel domain, and at the same time using on-line machine-learnt local spatial features to regulate the solution of the underlying inverse problem. Experimental results are encouraging and show the promise of the new approach in significantly improving the quality of DCT-coded images.

III. D3: Deep DualDomain Based Fast Restoration of JPEG-Compressed Images

Z. Wang, D. Liu, S. Chang, Q. Ling, Y. Yang and T. S. Huang, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2764-2772, doi: 10.1109/CVPR.2016.302.

Design a Deep Dual-Domain (D3) based fast restoration model to remove artifacts of JPEG compressed images. It leverages the large learning capacity of deep networks, as well as the problem-specific expertise that was hardly incorporated in the past design of deep architectures. For the latter, we take into consideration both the prior knowledge of the JPEG compression scheme, and the successful practice of the sparsity-based dual-domain approach. We further design the One-Step Sparse Inference (1-SI) module, as an efficient and lightweighted feed-forward approximation of sparse coding. Extensive experiments verify the superiority of the proposed D3 model over several state-of-the-art methods.

Specifically, our best model is capable of outperforming the latest deep model for around 1 dB in PSNR, and is 30 times faster.

IV. DeepSIC: Deep Semantic Image Compression

Luo, Sihui & Yang, Yezhou & Song, Mingli, International Conference on Neural Information Processing (ICONIP) (2018)

Incorporating semantic information into the codecs during image compression can significantly reduce the repetitive computation of fundamental semantic analysis (such as object recognition) in client-side applications. The same practice also enable the compressed code to carry the image semantic information during storage and transmission. In this paper, we propose a concept called Deep Semantic Image Compression (DeepSIC) and put forward two novel architectures that aim to reconstruct the compressed image and generate corresponding semantic representations at the same time. The first architecture performs semantic analysis in the encoding process by reserving a portion of the bits from the compressed code to store the semantic representations. The second performs semantic analysis in the decoding step with the feature maps that are embedded in the compressed code. In both architectures, the feature maps are shared by the compression and the semantic analytics modules. To validate our approaches, we conduct experiments on the publicly available benchmarking datasets and achieve promising results. We also provide a thorough analysis of the advantages and disadvantages of the proposed technique.

V. Image and Video Compression with Neural Networks: A Review

Ma, Siwei & Zhang, Xinfeng & Jia, Chuanmin & Zhao, Zhenghui & Wang, Shiqi & Wanga, Shanshe, IEEE Transactions on Circuits and Systems for Video Technology. PP. 1-1. 10.1109/TCSVT.2019.2910119.

In recent years, the image and video coding technologies have advanced by leaps and bounds. However, due to the popularization of image and video acquisition devices, the growth rate of image and video data is far beyond the improvement of the compression ratio. In particular, it has been widely recognized that there are increasing challenges of pursuing further coding performance improvement within the traditional hybrid coding framework. Deep convolution neural network (CNN) which makes the neural network resurge in recent years and has achieved great success in both artificial intelligent and signal processing fields, also provides a novel and promising solution for image and video compression. In this paper, we provide a systematic, comprehensive and up-to-date review of neural network based image and video compression techniques. The evolution and development of neural network based compression methodologies are introduced for images and video respectively. More specifically, the cutting-edge video coding techniques by leveraging deep learning and HEVC framework are presented and discussed, which promote the state-of-the-art video coding performance substantially. Moreover, the end-to-end image and video coding frameworks based on neural networks are also reviewed, revealing interesting explorations on next generation image and video coding frameworks/standards. And future trends are also envisioned, In particular, the joint compression on semantic and visual information is tentatively explored to formulate high efficiency signal representation structure for both human vision and machine vision, which are the two dominant signal receptor in the age of artificial intelligence.

III. EXPERIMENTAL RESULTS

A. Experimental Setup

It is a process of collecting and interpreting facts, identifying the problems, and decomposition of a system into its components. System analysis is conducted for the purpose of studying a system or its parts in order to identify its objectives. It is a problem solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose. Analysis specifies what the system should do. This chapter gives overview of architecture design, dataset for implementation, algorithm used and UML designs.

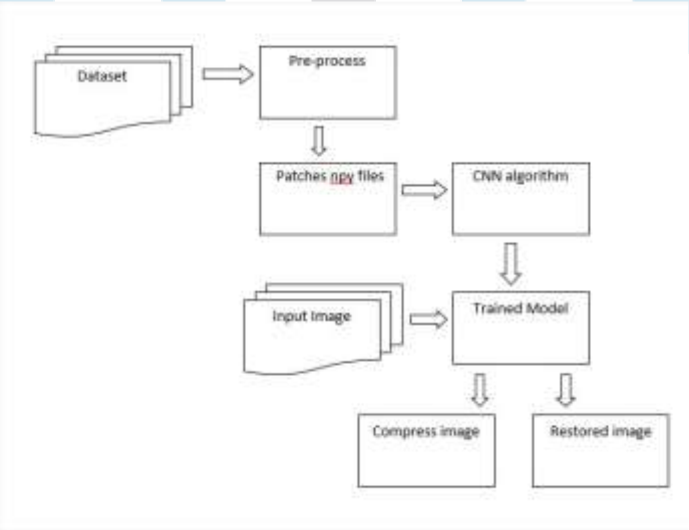

Fig 3: System Architecture

The design is a plan or drawing produced to show the look and function or workings of an object before it is made. Unified Modeling language (UML) is a standardized modeling language enabling developers to specify, visualize, construct and document artifacts of a software system. Thus, UML makes these artifacts scalable, secure and robust in execution.

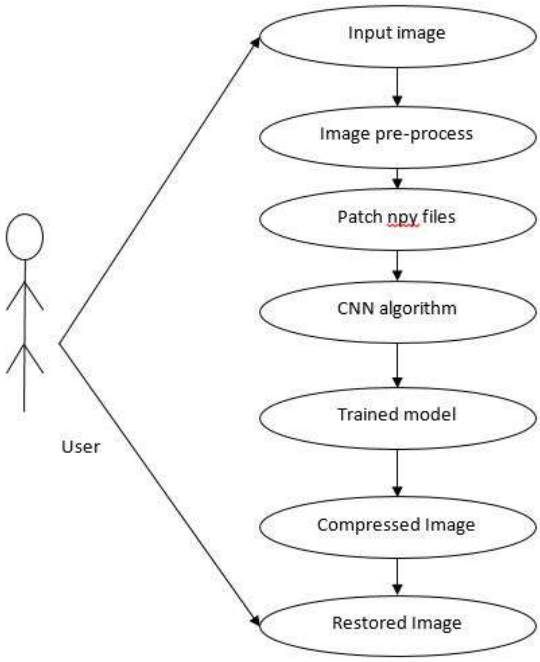

Fig 4: Use case Diagram

The above figure represent use case diagram of proposed system, where user inputs dataset, the algorithm work to generate the identified output. The actor and use case is represented. An eclipse shape represents the use case namely input image, pre-process, training, and output.

B. Efficiency Performance

The proposed work is implemented in Python 3 with libraries Keras, Tensorflow, OpenCV, and other mandatory libraries. This training dataset contains 1000 color images, each with dimension 256X256. Deep learning algorithm is applied Convolutional Neural Network. The result shows that the image is improved with good compression rate.

Clean Image

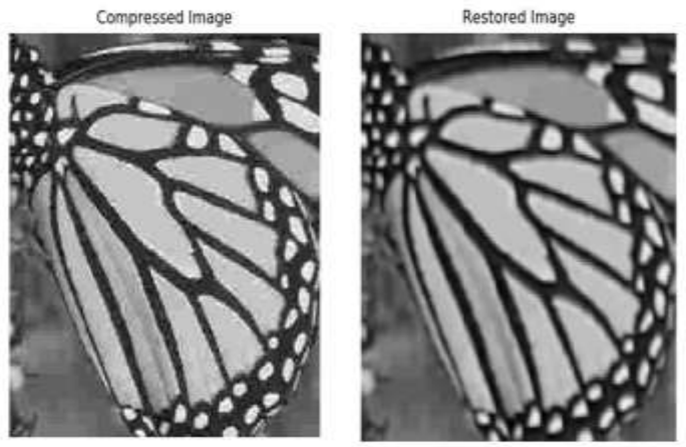

Fig 5: Resultant Diagram

IV. CONCLUSION AND FUTURE WORK

A deep convolutional neural network was proposed for image compression, where residual learning is adopted. The batch normalization and residual learning are integrated to speed up the training process as well as boost the performance. Moreover, we showed the feasibility to train a single DnCNN model to handle general image compression tasks of JPEG image with different quality factors.
In future, we are interested to address colored image for restorations.

REFERENCES

[1] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2005, pp. 60–65.
[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Transactions on Image Processing, vol. 16, no. 8, pp. 2080–2095, 2007.
[3] A. Buades, B. Coll, and J.-M. Morel, “Nonlocal image and movie denoising,” International Journal of Computer Vision, vol. 76, no. 2, pp. 123–139, 2008.
[4] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, “Non-local sparse models for image restoration,” in IEEE International Conference on Computer Vision, 2009, pp. 2272–2279.
[5] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng, “Patch group based nonlocal self-similarity prior learning for image denoising,” in International Conference on Computer Vision, 2015, pp. 244–252.
[6] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” IEEE Transactions on Image Processing, vol. 15, no. 12, pp. 3736–3745, 2006.
[7] W. Dong, L. Zhang, G. Shi, and X. Li, “Nonlocally centralized sparse representation for image restoration,” IEEE Transactions on Image Processing, vol. 22, no. 4, pp. 1620–1630, 2013.
[8] Z. Zha, X. Liu, X. Huang, X. Hong, H. Shi, Y. Xu, Q. Wang, L. Tang, and X. Zhang, “Analyzing the group sparsity based on the rank minimization methods,” arXiv preprint arXiv:1611.08983, 2016.
[9] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Physica D: Nonlinear Phenomena, vol. 60, no. 1, pp. 259–268, 1992.
[10] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, “An iterative regularization method for total variation-based image restoration,” Multiscale Modeling & Simulation, vol. 4, no. 2, pp. 460–489, 2005.
[11] Y. Weiss and W. T. Freeman, “What makes a good model of natural images?” in IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[12] X. Lan, S. Roth, D. Huttenlocher, and M. J. Black, “Efficient belief propagation with learned higher-order Markov random fields,” in European Conference on Computer Vision, 2006, pp. 269–282.