Unsupervised Image Compression via Cycle Consistency

Muwei He (2020.5.10)

1. Motivation

Images capture all aspects of our beautifully complex and diverse visual world. Yet, not every arrangement of pixel color values forms a real image. Most are illegible noise. This is the main insight behind lossless image compression. In fact, the Shannon source coding theorem directly links the likelihood of a group of pixels to be a real image, and our ability to compress that image.

Deep learning has been useful these years in so many research sub-fields of image processing while image compression is a typical one. Traditional image compression methods are mostly based on transform-coding such as BPG (G. J. Sullivan & Wiegand, Dec 2012) and JPEG (Taubman & Marcellin, 2013). With the recent advances in deep learning, neural networks have been applied to image compression with promising results.

The aim of compression is to save space as much as possible with minimal quality loss of source data after decompression. As for image compression problem, we think it is a **dual task of image super-resolution**. Inspired by CycleGAN (Jun-Yan Zhu & Efros, 2017), we propose a direct and simple deep learning based image compression algorithm via cycle consistency. Figure 1 shows the rough description of **cycle consistency**. Explain in detail, in image compression, if we find a encoder which can be a deep neural network for compression, then its "symmetrical" decoder should reconstruct the compressed image into source image. The compressed image don't need to be a seemed meaningful image, which could be a two-dim integer vector, one-dim float vector or even one-dim binary vector. Whatever how simple the compressed code is, as long as the fully trained decoder could reconstruct the compressed code into source image, the encoder-decoder is successful. Our proposed algorithm is also inspired by (Caglar Aytekin & Hannuksela, 24 May 2019) with some additional design and details which will be described in **section 2**.

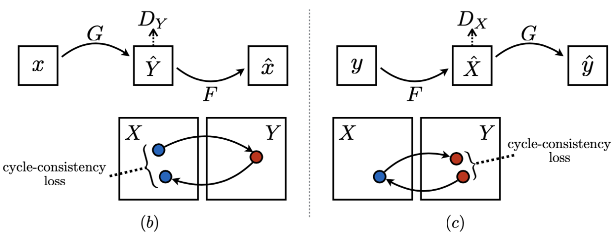

Figure 1. Conception of Cycle Consistency

2. Methodology

As mentioned above, we want to train a encoder-decoder, of which the input is a single image I, with which we can get compressed code of a given image called $\sigma(I)$. Then the decoder could reconstruct the image I' from code $\sigma(I)$ as similar as the source image I. However, if we directly force the raw pixel matrix of I and I' to be same, the network will be hard to converge from experience so we want the compressed code of reconstructed I' $\sigma(I')$ to be similar with $\sigma(I)$ which is more reasonable and feasible. See figure 2., our proposed algorithm is shown. It can be described as:

$$Encoder(I) = Encoder(Decoder(Encoder(I)))$$

by which the "cycle consistency" is shown. See figure2, we design a prepositive down-sampling block with multiple residual blocks as encoder and its dual or called symmetrical structure, rear up-sampling block with multiple residual blocks as decoder. The encoder and decoder should share parameters of each layer correspondingly. The form or the size of compressed code could be modified by changing the last layer of encoder. In addition, the compressed code should be as simple or say "sparse" as possible.

We designed two loss functions to supervise the training process. The first one cyclic consistency loss which is:

$$L_{cons} = \sum_{a}^{M} MSE(\sigma(I_a), \sigma(I'_a)) = \sum_{a}^{M} \frac{1}{N} \sum_{i}^{N} (I_a(i) - I'_a(i))^2 \tag{1}$$

The M is the total number of training images and N is the dimension of compressed code $\sigma(image)$. By minimizing this loss function, the compressed code of source image I and reconstructed image I' will be close. Since we hope the compressed code be more sparse with least zero-value, we need another loss function called compression loss:

$$L_{comp} = \frac{|x|}{||x||} + \alpha \frac{||x||^2}{|x|} \tag{2}$$

This is adopted from the work (Hoyer, 2004) and is a measure of the sparsity in a signal. We call this sparsity term of the compression objective. The sparsity term is independent of the values of non-zeros in the signal. For example a vector $[0, 0, 500, 500]$ and $[0, 0, 0.1, 0.1]$ would have exactly the same sparsity value and large values are not penalized. However, it is usually a good practice to have reasonably small values in machine learning literature to

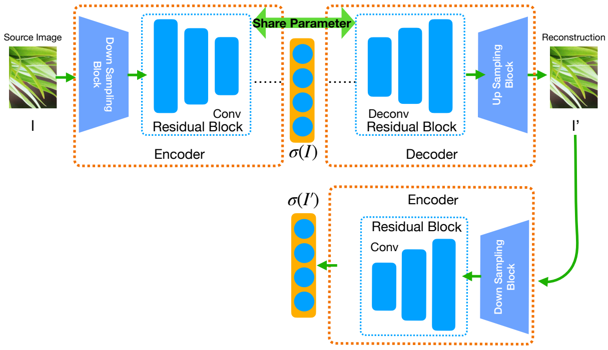

Figure 2. Algorithm Visualization

avoid exploding gradients and also to act as a regularization. Because of this, we add another factor to the compression loss which favors small non-zero values in a signal – this is the second part: $\frac{||x||^2}{|x|}$. The weight α in Eq. 2, acts as a regularizer between the sparsity term and the squeezing term (Caglar Aytekin & Hannuksela, 24 May 2019). Then we could combine them together as total training loss:

$$L = \lambda_{cons} L_{cons} + \lambda_{comp} L_{comp} \tag{3}$$

while the λ_{cons} and λ_{comp} are both real number from 0.1 to 1 and $\lambda_{cons} + \lambda_{comp} = 1$.

3. Dataset and Evaluation

We gonna evaluate our algorithm on OpenImages (Krasin, 2017) which consists of high resolution images. We plan to use 2 different versions of Open Images, JPEG 3 and PNG. The quantity of training images will be decided before experiments startting. The evaluation matrix we focus on are **Compression Rate, Processing Speed per hundred images and Reconstruction Quality**. We measure compression rate by bits per subpixel (bpsp) and run-time on single GPU Nvidia 1080 Ti. The reconstruction quality is hard to quantized and can only evaluated by human.

References

Caglar Aytekin, Francesco Cricri, Antti Hallapuro Jani Lainema Emre Aksu and Hannuksela, Miska. A compression objective and a cycle loss for neural image compression. *arXiv:1905.10371v1 [eess.IV]*, 24 May 2019.

G. J. Sullivan, J. R. Ohm, W. J. Han and Wiegand, T. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on Circuits and Systems for Video Technology, 22(12)*, pp. 1649–1668, Dec 2012.

Hoyer, O. P. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, pp. 1457–1469, 2004.

Jun-Yan Zhu, Taesung Park, Phillip Isola and Efros, Alexei A. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.

Krasin, I., Duerig T. Alldrin N. Ferrari V. Abu-El-Haija S. Kuznetsova A. Rom H. Uijlings J. Popov S. Veit A. et al. : Open images: A public dataset for large-scale multi-label and multi-class image classification. *Dataset available from https://github. com/openimages*, 2017.

Taubman, D. and Marcellin, M. Jpeg2000 image compression fundamentals, standards and practice. *Springer Publishing Company, Incorporated*, 2013.