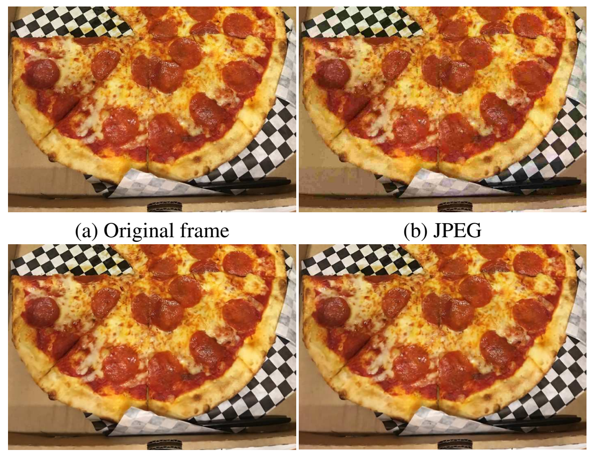
Efficient Learning Based Sub-pixel Image Compression

Chunlei Cai, Guo Lu, Qiang Hu, Li Chen, and Zhiyong Gao *

Shanghai Jiao Tong University
{caichunlei, luguo2014, hq2902108007, hilichen, zhiyong.gao}@sjtu.edu.cn

Abstract

In this paper, we propose an efficient learning based sub-pixel image compression algorithm. Our framework builds upon the previous variational auto-encoder architecture and reduces the computational complexity significantly. Specifically, we propose an end-to-end optimized image compression framework to utilize the powerful non-linear representation ability of neural networks. This framework follows the widely used variational auto-encoder architecture and is optimized based on the rate-distortion balance. More importantly, a sub-pixel image compression framework is exploited to reduce the spatial resolution of image and improve the inference speed. Experimental results demonstrate the effectiveness of our method. Compared with the baseline algorithm, our encoder is 2 times faster with negligible performance decrease. The decoding speed of our method for the CLIC dataset is 1.85 fps on GTX 1080Ti, which makes our codec one of the fastest learning based image compression algorithm.



(a) Original frame (b) JPEG

(c) Our *slowdic* (d) Ours *fastdic*

Figure 1: Visual quality of the reconstructed images from different image compression systems. (a) Original (b) JPEG (c) Ours *slowdic* (d) Ours *fastdic*.

1. Introduction

Image compression technique is widely used to reduce the size of the storage for image. Traditional image compression algorithms (such as JPEG [20] or JPEG2000 [16]) utilize hand-crafted techniques (such as discrete cosine transform (DCT)) to reduce the spatial redundancy in the image. However, the traditional image compression pipeline is not end-to-end optimized, which means the performance could be further improved by feasiable approaches. In addition, linear transform (e.g, DCT) used in the traditional image codecs can also be replaced by the powerful non-linear transform (such as deep neural networks (DNN)).

To this end, a lot of DNN based auto-encoder for image compression [18, 7, 19, 4, 8, 9, 17, 10, 15, 5, 13] have been proposed for image compression. The learning based approaches have achieved comparable or even better performance than the traditional image codecs like JPEG [20], JPEG2000 [16] or BPG [2]. One possible explanation is that the DNN based image compression methods can exploit large scale end-to-end training and highly non-linear transform, which are not used in the traditional approaches.

However, one disadvantage of current learning based approaches is the huge computational complexity. Compared with traditional image codecs such as BPG [2] or WebP [3], recent proposed learning based approaches need much more computing resource. For example, the parameters and FLOPs of recent state-of-the-art image compression algorithm [13] are 7.86M and 295 GFlops, and the corresponding speed is 0.85fps on the CLIC dataset [1]. One possible reason is that the current learning based approaches perform transform/convolution based on the *full* resolution image, which increases the computational complexity, especifically for the image with large resolution.

In this paper, we proposed an efficient learning based

*Corresponding author

sub-pixel image compression algorithm. Our framework is based on the variational auto-encoder compression architecture [7, 8]. We analyze the computational bottleneck of current learning based approaches and proposed a straightforward approach to reduce the computational burden. Specifically, we use the sub-pixel layer in the VAE architecture and builds an efficient image compression framework. Our approaches can reduce the huge computational complexity for the full resolution image/feature and achieve efficient learning based image compression. Experimental results demonstrate that the proposed approach is 2 times faster the baseline algorithm with negligible performance decrease.

2. Related Work

2.1. Traditional Image Compression

Traditional image compression algorithms [20, 16, 2, 3] have been widely studied in the past decades. Traditional pipeline utilizes the transform techniques to map the pixel values to the corresponding latent space, which is more compact. For example, the JPEG algorithm linearly maps the pixels to the latent space by using DCT, and quantizes the corresponding coefficients before entropy coding[20]. Although the traditional image compression algorithms are efficient by using these hand-crafted techniques, it is possible to further improve the coding efficiency by building an end-to-end optimized framework for image compression.

2.2. Learning Based Image Compression

Recently, DNN based image compression methods [18, 19, 7, 8, 17, 4, 10, 15, 12, 5, 13] have achieved surprising coding efficiency. These methods can be roughly categorized into two kinds. First, in [18, 19, 9], recurrent neural networks (RNNs) are utilized to build a progressive image compression scheme. Their approach is further improved by using other RNN architectures, learning based entropy model, priming and spatial adaptive bitrate allocation [19, 9]. These RNN based approaches can generate different target bitrates through a single model, which is beneficial for the rate control scheme. However, the RNN approaches utilizes the convolutional LSTM/GRU module, which decreases the inference speed. In addition, these RNN based approaches only optimizes the network by minizing the distortion term instead of considering the rate-distortion trade-off.

On the other hand, other methods try to employ the CNNs to design an auto-encoder style network for image compression [7, 8, 17, 10, 13]. More importantly, rate-distortion optimization technique was adopted in [7, 8, 17, 10] for higher compression efficiency by introducing the number of bits in the optimization procedure. To estimate the bit rates, context models are learned for the adaptive arithmetic coding method in [15, 10, 12], while non-adaptive arithmetic coding is used in [7, 17]. In [7], generalized divisive normalization (GDN) is utilized to normalize the image with better compression efficiency. In [15], a multi-scale image decomposition scheme is employed to exploit the scale information and the adversarial training is also utilized for high quality reconstruct images. In [10, 12], importance map is generated to guide the coding procedure for the latent representation for higher coding efficiency. In addition, intra prediction techniques are also utilized in [14, 6] through the image inpainting.

Recently, learning based video compression [21, 11] have attracted more and more attention. Lu *et al.* proposed an end-to-end optimized video compression framework by compressing both the motion information and residual information.

3. Proposed Method

3.1. Overview of the Basic Image Compression Model

Our proposed sub-pixel image compression architecture is illustrated in the Figure 1.

For the image compression task, the input image x is mapped to the latent space through the encoder network g_a and the corresponding latent representation is y. After the quantization step, we can get the quantized representation $\hat{y}$. Then the reconstructed image $\hat{x}$ is obtained based on the decoder decoder g_s. In order to improve the coding efficiency, we follows the design in [8] and employ the hyperprior network to reduce the spatial redundancy in the latent representation $\hat{y}$. Specifically, the corresponding hyperprior encoder/decoder network (h_a and h_s) are employed to get the spatial relationship in the latent space and the reconstructed $\hat{\sigma}$ is utilized to model the distribution of latent representation $\hat{y}$. Please refer to [8] for more details.

3.2. Proposed Sub-pixel Image Compression

Although current learning based image compression algorithms (such as [13]) achieve state-of-the-art coding efficiency, these methods requires a lot of computational resource and the inference speed is slow. One possible explanation is that the previous image compression algorithms perform transform on the *full* resolution, which may bring the huge computational complexity.

In this paper, we introduce the sub-pixel based image compression scheme. Let r denote the scale of the sub-pixel layer. The size of image x_t is $W \times H \times Cd$. Before feeding the input image x into the encoder network g_a, we first perform sub-pixel transform and get the corresponding low resolution feature x_d with the size of $W/r \times H/r \times r^2C$. After the decoder network g_s in our sub-pixel approach, the reconstruct image with size of $W/r \times H/r \times r^2C$ will be utilized to get the full resolution image $\hat{x}_t$. Based on our

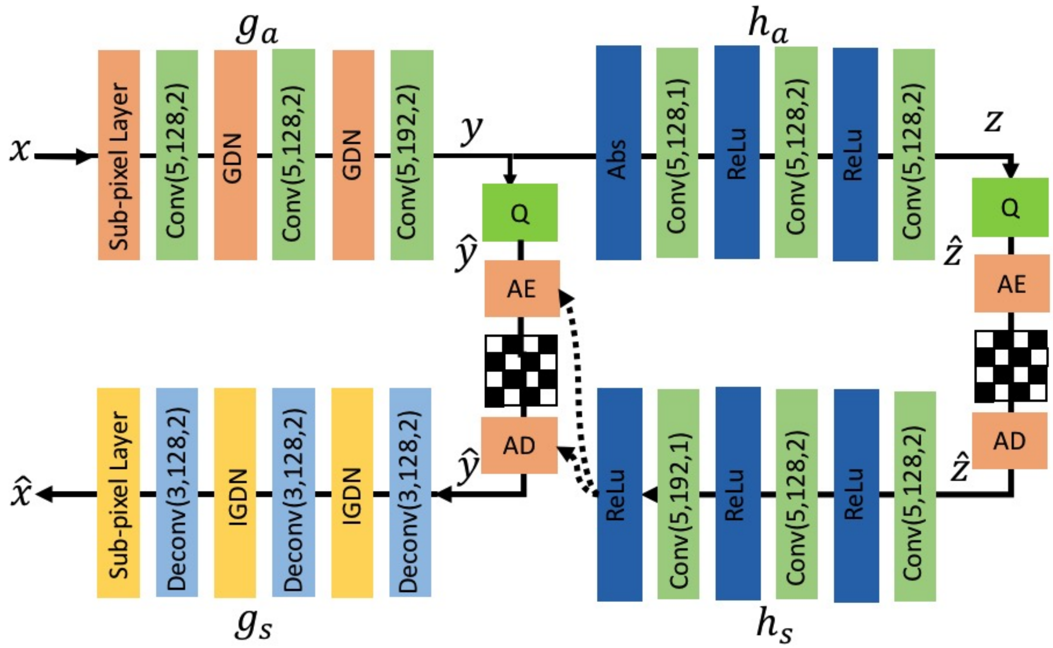


Figure 2: Overall architecture of the proposed sub-pixel image compression.

proposed method, we can reduce the computational complexity for the feature map with large resolution. Our approach can also be easily extended to other learning based image compression algorithm.

3.3. Training Procedure

Our proposed sub-pixel image compression framework is optimized by minimizing the rate-distortion trade-off. The loss function $\mathcal{L}$ is formulated as follows,

$$\mathcal{L} = \lambda * D + R \tag{1}$$

where D is the distortion function and R is the bitrate. In our implementation, we use mean square error to measure the distortion. The R is the estimated bitrate and we use the network architecture in [8] to calculate the bits. λ is the trade-off parameter.

In addition, in order to train the network in an end-to-end manner, it is critical to approximate the quantization operation in the whole network. In our approach, we replace the quantization step by adding uniform noise in the training stage. In the reference stage, we use the rounding operation.

4. Experiments

4.1. Experimental Setup

In this paper, we use the CLIC dataset to train our framework. The learning rate is 1e-4 and the whole network is optimized for 2M iterations. In order to invalid the effectiveness of our approach, we perform experiments on the widely used the CLIC validation dataset [1].

Implementation Details Our framework is implemented based on the Tensorflow platform and it takes 48 hours to train our network on GTX 1080Ti.

4.2. Experimental Results

We submit two codecs: *fastdic* and *slowdic*. *slowdic* is the implementation of the method in [13]. *fastdic* is the proposed sub-pixel framework based on the method in [8]. In this subsection, we first provide the our experimental results on the CILC 2019 *low-rate* track. It is obvious that our sub-pixel framework (*fastdic*) is the *fastest* among the top-10 codecs when measured by MS-SSIM. The total decoding time of ours is *160147 ms* in the leaderboard.

4.3. Ablation Study and Model Analysis

In our proposed framework, we use the sub-pixel scheme for image compression. In this subsection, we analyse the computational complexity between different codecs. Our approach is based on the VAE codec [8]. The inference speed of the baseline algorithm [8] is 0.85 fps while ours is 2 times faster.

5. Conclusion

In this paper, we proposed the a learning based sub-pixel image compression algorithm. Our framework can reduce the computational complexity that introduced by the feature map with large spatial resolution. Our framework is straightforward and effective, and can be easily entended to other image compression algorithms.

References

[1] Clic 2019. http://www.compression.cc/. Accessed: 2018-5-5. 1, 3

[2] F. bellard, bpg image format. http://bellard.org/bpg/. Accessed: 2018-10-30. 1, 2

[3] Webp. https://developers.google.com/speed/webp/. Accessed: 2018-10-30. 1, 2

[4] E. Agustsson, F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini, and L. V. Gool. Soft-to-hard vector quantization for end-to-end learning compressible representations. In *NIPS*, pages 1141–1151, 2017. 1, 2

[5] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool. Generative adversarial networks for extreme learned image compression. *arXiv preprint arXiv:1804.02958*, 2018. 1, 2

[6] M. H. Baig, V. Koltun, and L. Torresani. Learning to inpaint for image compression. In *NIPS*, pages 1246–1255, 2017. 2

[7] J. Ballé, V. Laparra, and E. P. Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016. 1, 2

[8] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436*, 2018. 1, 2, 3

[9] N. Johnston, D. Vincent, D. Minnen, M. Covell, S. Singh, T. Chinen, S. Jin Hwang, J. Shor, and G. Toderici. Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks. In *CVPR*, June 2018. 1, 2

[10] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang. Learning convolutional networks for content-weighted image compression. In *CVPR*, June 2018. 1, 2

[11] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao. Dvc: An end-to-end deep video compression framework. *arXiv preprint arXiv:1812.00101*, 2018. 2

[12] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool. Conditional probability models for deep image compression. In *CVPR*, number 2, page 3, 2018. 2

[13] D. Minnen, J. Ballé, and G. D. Toderici. Joint autoregressive and hierarchical priors for learned image compression. In *Advances in Neural Information Processing Systems*, pages 10771–10780, 2018. 1, 2, 3

[14] D. Minnen, G. Toderici, M. Covell, T. Chinen, N. Johnston, J. Shor, S. J. Hwang, D. Vincent, and S. Singh. Spatially adaptive image compression using a tiled deep network. In *ICIP*, pages 2796–2800. IEEE, 2017. 2

[15] O. Rippel and L. Bourdev. Real-time adaptive image compression. In *ICML*, 2017. 1, 2

[16] A. Skodras, C. Christopoulos, and T. Ebrahimi. The jpeg 2000 still image compression standard. *IEEE Signal Processing Magazine*, 18(5):36–58, 2001. 1, 2

[17] L. Theis, W. Shi, A. Cunningham, and F. Huszár. Lossy image compression with compressive autoencoders. *arXiv preprint arXiv:1703.00395*, 2017. 1, 2

[18] G. Toderici, S. M. O'Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, and R. Sukthankar. Variable rate image compression with recurrent neural networks. *arXiv preprint arXiv:1511.06085*, 2015. 1, 2

[19] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell. Full resolution image compression with recurrent neural networks. In *CVPR*, pages 5435–5443, 2017. 1, 2

[20] G. K. Wallace. The jpeg still picture compression standard. *IEEE Transactions on Consumer Electronics*, 38(1):xviii–xxxiv, 1992. 1, 2

[21] C.-Y. Wu, N. Singhal, and P. Krahenbuhl. Video compression through image interpolation. In *ECCV*, September 2018. 2