It is known that, with a given quality, a recurrent neural network (RNN) can be trained to achieve that quality. [17] Moreover, image compressions via autoencoders [9, 13, 18] are also becoming more and more common, so are their combinations [6]. In this paper, three methods combining convolutional neural network (CNN), RNN, autoencoders, and Generative Adversarial Networks (GANs) are proposed, trained, and tested. The commonly used performances measurements of image compression algorithms Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Multi-Scale Structural Similarity (MS-SSIM) [19] are used to compare the result side by side with that of JPEG.

Peak Signal to Noise Ratio (PSNR)

MS-SSIM [19]

[17] showed that it is possible to train a single recurrent neural network and achieve better than state of the art compression rates for a given quality regardless of the input image, but was limited to 32×32 images.

Autoencoders have been used to reduce the dimensionality of images [9], convert images to compressed binary codes for retrieval [13], and to extract compact visual representations that can be used in other applications [18].

variational (recurrent) autoencoders have been directly applied to the problem of compression [6] (with results on images of size up to 64×64 pixels), while non-variational recurrent neural networks were used to implement variable rate encoding [17].

**Introduction**

[17] showed that it is possible to train a single recurrent neural network and achieve better than state of the art compression rates for a given quality regardless of the input image, but was limited to 32×32 images. In that work, no effort was made to capture the long-range dependencies between image patches.

Our goal is to provide a neural network which is competitive across compression rates on images of arbitrary size. There are two possible ways to achieve this: 1) design a stronger patch-based residual encoder; and 2) design an entropy coder that is able to capture long-term dependencies between patches in the image.

In this paper, we address both problems and combine the two possible ways to improve compression rates for a given quality. In order to measure how well our architectures are doing (i.e., “quality”), we cannot rely on typical metrics such as Peak Signal to Noise Ratio (PSNR), or Lp differences between compressed and reference images because the human visual system is more sensitive to certain types of distortions than others. This idea was exploited in lossy image compression methods such as JPEG. In order to be able to measure such differences, we need to use a human visual-system-inspired measure which, ideally should correlate with how humans perceive image differences. Moreover, if such a

metric existed, and were differentiable, we could directly optimize for it. Unfortunately, in the literature there is a wide variety of metrics of varying quality, most of which are non-differentiable. For evaluation purposes, we selected two commonly used metrics, PSNR-HVS [7] and MS-SSIM [19], as discussed in Section 4.

**1.1**

Autoencoders have been used to reduce the dimensionality of images [9], convert images to compressed binary codes for retrieval [13], and to extract compact visual representations that can be used in other applications [18].

More recently, variational (recurrent) autoencoders have been directly applied to the problem of compression [6] (with results on images of size up to 64×64 pixels), while non-variational recurrent neural networks were used to implement variablerate encoding [17].

Most image compression neural networks use a fixed compression rate based on the size of a bottleneck layer [2]. This work extends previous methods by supporting variable rate compression while maintaining high compression rates beyond thumbnail-sized images.

[2] J. Ballé, V. Laparra, and E. P. Simoncelli. End-to-end optimization of nonlinear transform codes for perceptual quality. In Picture Coding Symposium, 2016. 1

[6] K. Gregor, F. Besse, D. Jimenez Rezende, I. Danihelka, and D. Wierstra. Towards Conceptual Compression. ArXiv eprints, 2016. 1

[9] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786):504–507, 2006. 1

[13] A. Krizhevsky and G. E. Hinton. Using very deep autoencoders for content-based image retrieval. In European Symposium on Artificial Neural Networks, 2011. 1

[17] 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. ICLR 2016, 2016. 1, 2, 3, 5, 7

[18] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. Journal of Machine Learning Research, 2012. 1

[19] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on, volume 2, pages 1398–1402. Ieee, 2003.