# Efficient Image Compression

**James Camacho**
MIT / AI
jamesc03@mit.edu

**Linda He**
Harvard / Applied Mathematics
lindahe@college.harvard.edu

## Abstract

With high-dimensional spaces such as images or video, perfect communication becomes prohibitively expensive. A lossy, compressed version is cheaper to transmit and often good enough for most purposes. Traditional algorithms such as JPEG use a Fourier transform to pick out the most important features for transmission. In this paper, we explore using auto- and raster-encoders to automatically and efficiently compress images instead. We find they significantly outperform JPEG on the MNIST dataset, and discuss potential future improvements to their speed and cost via reinforcement learning.

## 1 Introduction

Over 75% of internet traffic comes in the form of video (Cisco, 2018), and YouTube alone nets $30bn from their distribution (Alphabet Inc., 2024). At these massive scales, every extra bit of compression is important. In this paper, we explore more optimal compression with the use of deep learning methods, including one adapted from language models.

Image and text have often been treated as separate domains, but there is much that can be applied between the two. Inspired by quantization and fractalization from image compression, Witten *et al.* proposed several lossy text encoding schemes as far back as 1992 (Witten et al., 2000). Recent advances in image generation such as denoising models (Ho et al., 2020) have similarly found applications in text generation by directly translating pixels into language tokens (Kou et al., 2024). Going the other direction, the popular Transformer architecture of language models have been applied to vision (Dosovitskiy et al., 2021), and has seen state-of-the-art success in video production (Liu et al., 2024).

As we will see in 3.1, the Vision Transformer algorithm bears a striking resemblance to JPEG. This is no accident. Generation is the inverse of compression, and more generally "being able to compress well is closely related to acting intelligently" (Hutter, 2020). Unfortunately, this enforces a tradeoff between the algorithm's speed and size. For example, an unintelligent compressor may store images verbatim, or a generator may simply sample from its training dataset. Smarter algorithms will do much better, but require more computation.

This paper will focus on the tradeoff between image quality and compression length, but we will end with several suggestions for improving the running time via reinforcement learning.

## 2 Related Work

## 3 Methods

### 3.1 JPEG

These vision transformers bear a striking resemblance to the JPEG algorithm

### 3.2 Auto-Encoders

Reference the paper about how CNNs find features like stripes.

### 3.3 Raster-Encoders

## 4 Discussion

Testing this: This is some text with a citation (Lazaridou et al., 2020).

## Acknowledgments

## References

Alphabet Inc. 2024. Alphabet 2024 q1 earnings report.

Cisco. 2018. Global device growth and traffic profiles.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An image

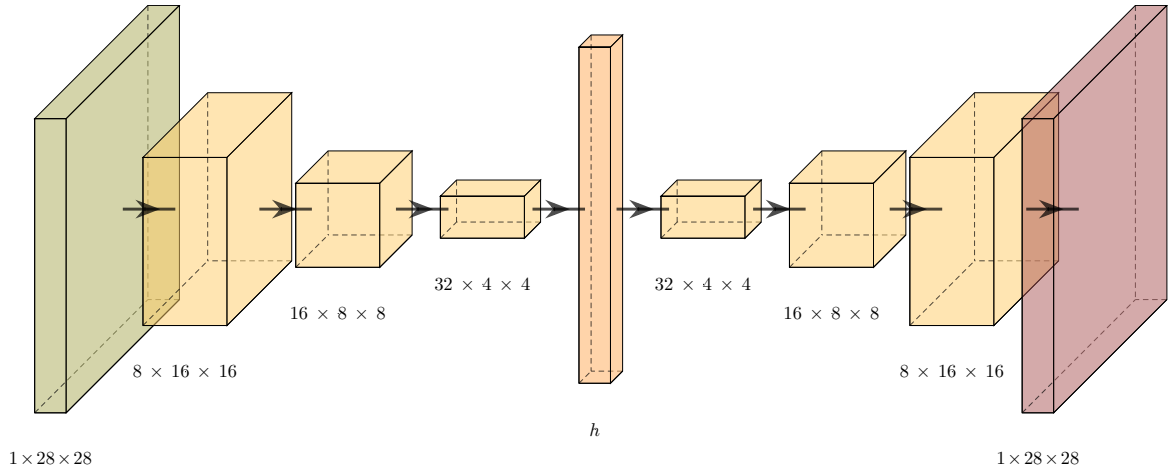


Figure 1: Auto-encoder network.

is worth 16x16 words: Transformers for image recognition at scale. *Preprint*, arXiv:2010.11929.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Preprint*, arXiv:2006.11239.

Marcus Hutter. 2020. Universal artificial intelligence, aixi, and agi. Video interview.

Siqi Kou, Lanxiang Hu, Zhezhi He, Zhijie Deng, and Hao Zhang. 2024. Cllms: Consistency large language models. *Preprint*, arXiv:2403.00835.

Angeliki Lazaridou, Anna Potapenko, and Olivier Tieleman. 2020. Multi-agent communication meets natural language: Synergies between functional and structural language learning. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7663–7674, Online. Association for Computational Linguistics.

Yixin Liu, Kai Zhang, Yuan Li, Zhiling Yan, Chujie Gao, Ruoxi Chen, Zhengqing Yuan, Yue Huang, Hanchi Sun, Jianfeng Gao, Lifang He, and Lichao Sun. 2024. Sora: A review on background, technology, limitations, and opportunities of large vision models. *Preprint*, arXiv:2402.17177.

Ian Witten, Timothy Bell, Alistair Moffat, Craig Nevill-Manning, Cowlemon Tony, and Harold Thimbleby. 2000. Semantic and generative models for lossy text compression. *The Computer Journal*, 37.

## A Example Appendix

This is an appendix.

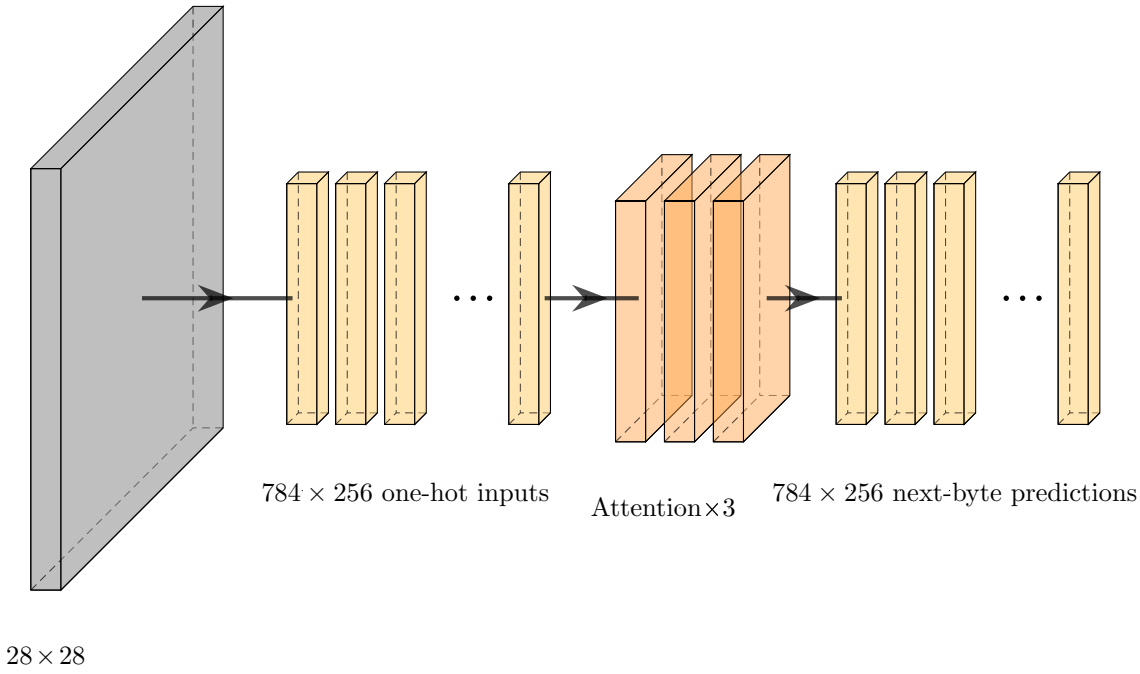


Figure 2: Raster-encoder network.