Full Resolution Image Compression with Recurrent Neural Networks

George Toderici
Google Research
gtoderici@google.com

Damien Vincent
damienv@google.com

Nick Johnston
nickj@google.com

Sung Jin Hwang
sjhwang@google.com

David Minnen
dminnen@google.com

Joel Shor
joelshor@google.com

Michele Covell
covell@google.com

A brief review of similar attempts ..

- In the paper "Using very deep autoencoders for content-based image retrieval" A. Krizhevsky and Geoffrey Hinton proposed autoencoder can be used to map images to binary codes.
- The paper "Extracting and Composing Robust Features with Denoising Autoencoders" proposes use of initial unsupervised step in denoising autoencoder.
- In march 2016, a team at google proposed LSTMs for autoencoder. The paper "Variable rate image compression with recurrent neural networks" describes how thumbnails generation can be accomplished.

Introduction

- Proposed neural network consists of RNN based encoder and decoder and a binarizer.
- This framework provides competitive compression rates on images of arbitrary size.
- Binarizer is a neural network used for entropy coding.
- The network is evaluated using : PSNR-HVS and MS-SSIM metrics.
- This architecture outperforms JPEG on kodak image dataset.

Network architecture

- There are two ways in which an image can be reconstructed. - one shot reconstruction and additive reconstruction.
- Single iteration of network can be represented as :

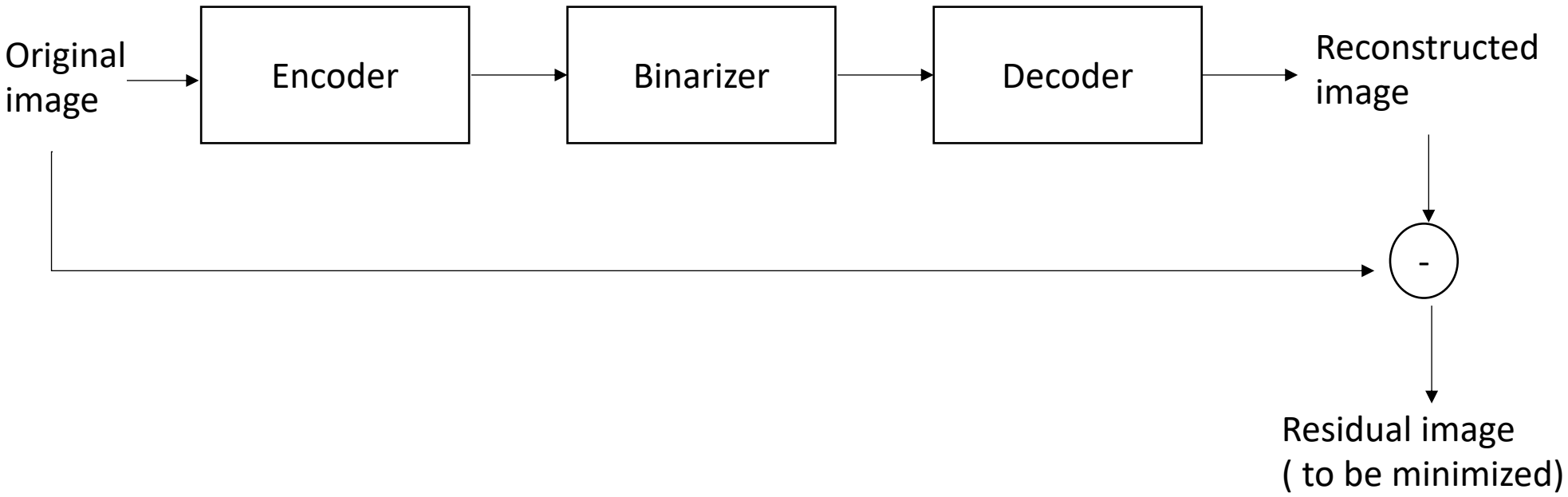
$$b_t = B(E_t(r_{t-1})), \quad \hat{x}_t = D_t(b_t) + \gamma \hat{x}_{t-1},$$
$$r_t = x - \hat{x}_t, \quad r_0 = x, \quad \hat{x}_0 = 0$$

Here γ = 0. represents one-shot reconstruction while γ = 1 represents additive reconstruction.

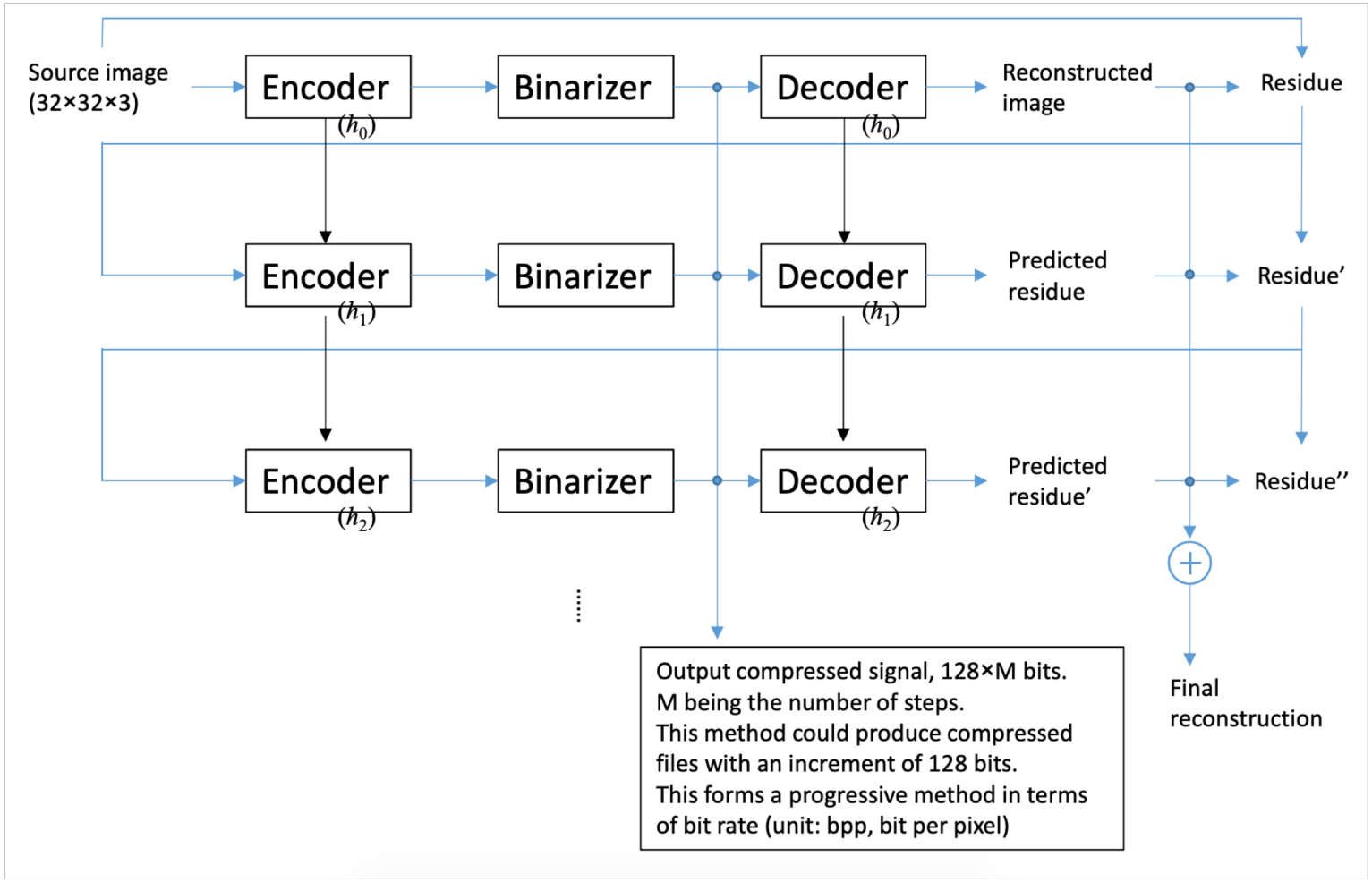
$$b_t = \{-1,1\}^{\mathrm{m}}$$

m is number of bits produced after each iteration.

Network Architecture



Unrolled model



Single iteration of architecture

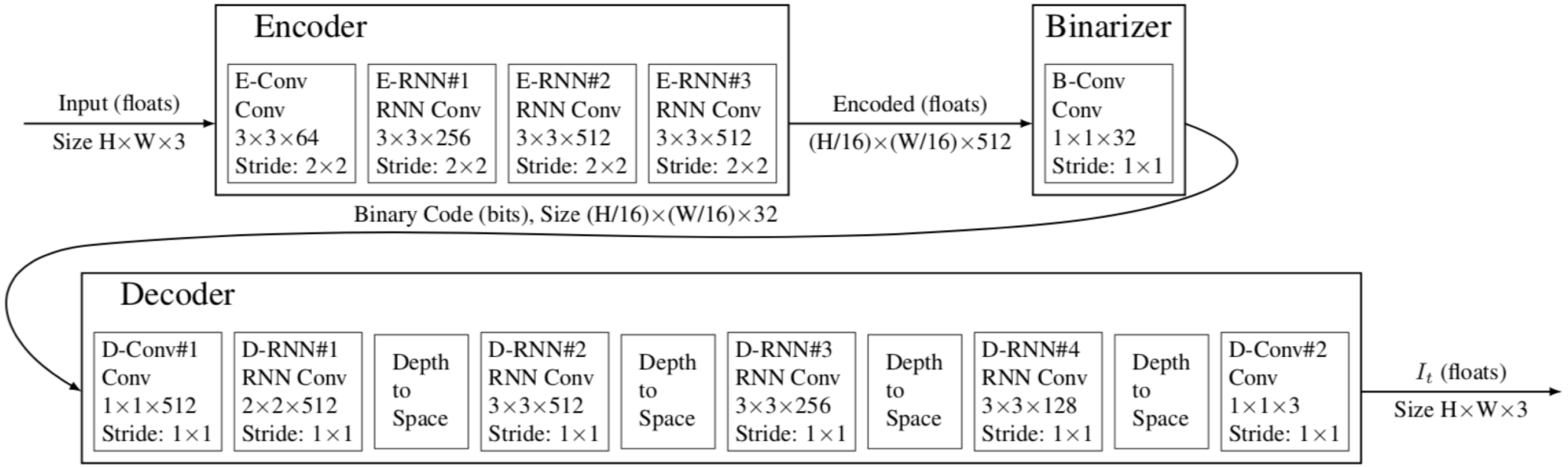
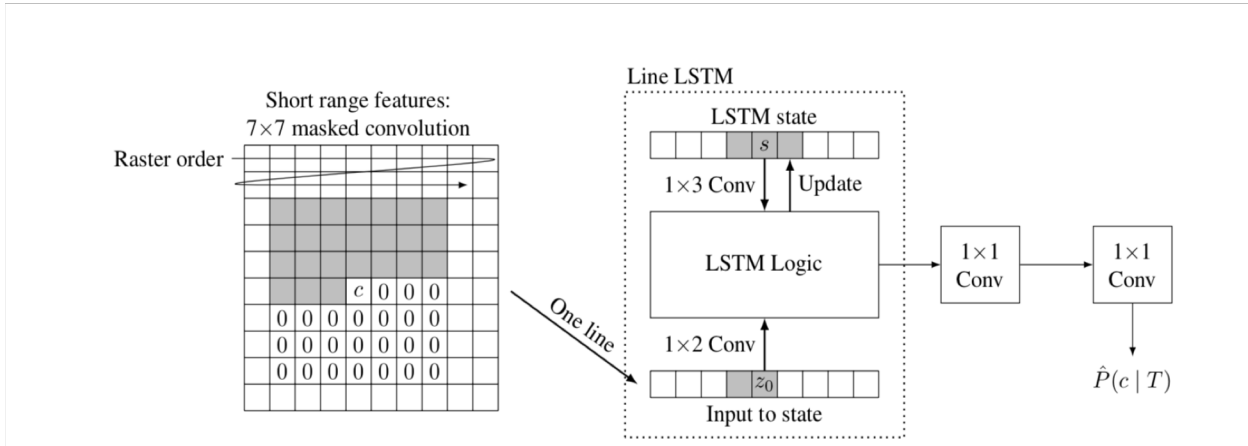


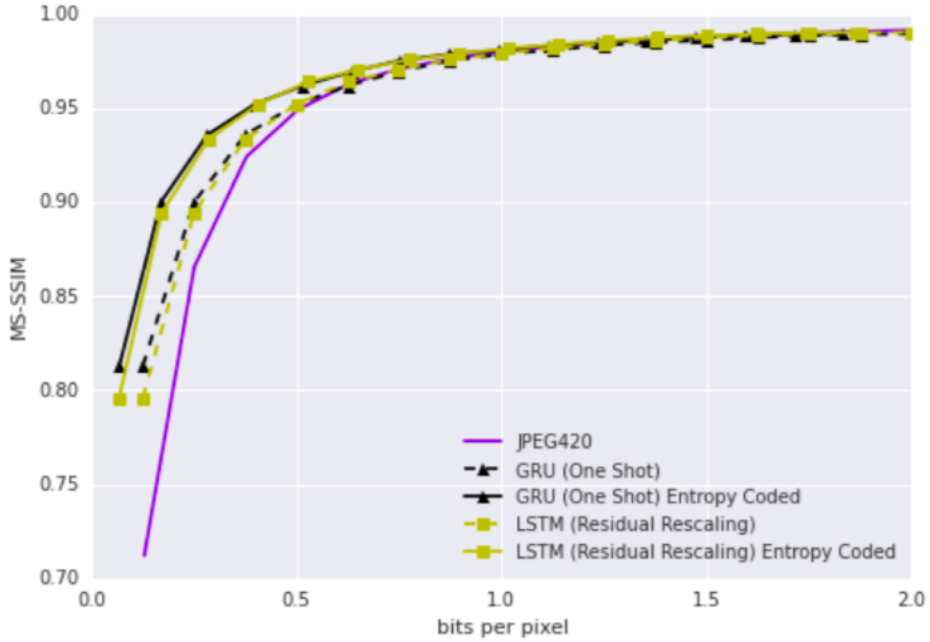
Figure 1. A single iteration of our shared RNN architecture.

How conv + RNN cell works?

- Convolution followed by LSTM cell.
- Divide tensor in small chunks.
- Apply LSTM on individual chunks.
- Return value of hidden state and c.



Results



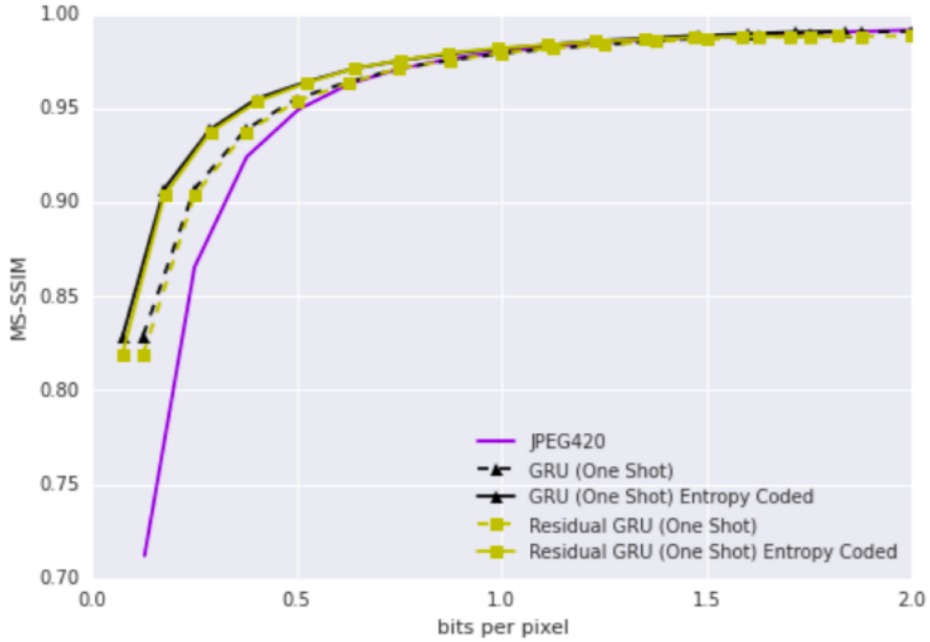


Figure 5. Rate distortion curve on the Kodak dataset given as MS-SSIM vs. bit per pixel (bpp). Dotted lines: before entropy coding, Plain lines: after entropy coding. Left: Two top performing models trained on the 32x32 dataset. Right: Two top performing models trained on the High Entropy dataset.