VRIC: Variable Rate Image Compression with Recurrent Neural Networks

Coding of Audiovisual Contents

Miquel Oller Oliveras & Alvaro Scherk Fontanals

Barcelona, December 18, 2017

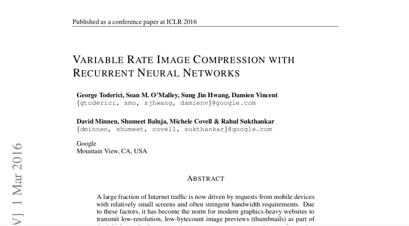
UPC
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

telecom BCN
Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

Table of Contents:

VRIC: An overview of the proposal

An overview of the proposal: Abstract[1]



Published as a conference paper at ICLR 2016

VARIABLE RATE IMAGE COMPRESSION WITH RECURRENT NEURAL NETWORKS

George Toderici, Sean M. O'Malley, Sung Jin Hwang, Damien Vincent
{gtoderici, smo, sjhwang, damienv}@google.com

David Minnen, Shumeet Baluja, Michele Covell & Rahul Sukthankar
{dminnen, shumeet, covell, sukthankar}@google.com

Google
Mountain View, CA, USA

ABSTRACT

A large fraction of Internet traffic is now driven by requests from mobile devices with relatively small screens and often stringent bandwidth requirements. Due to these factors, it has become the norm for modern graphics-heavy websites to transmit low-resolution, low-bytecount image previews (thumbnails) as part of

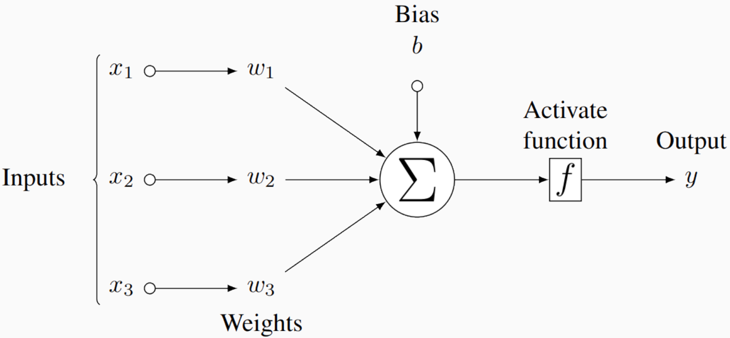
- Low-resolution thumbnail images
- Variable fixed encoding size
- Standards do not perform well with low-dimentional images
- Fresh air into image encoding world
- Trained end-to-end

[1] Proposal by George Toderici et al. All from the Google, Mountain View, CA, USA. Published in ICLR 2016

Neural Networks Overview

Neural Networks Overview: General Knowledge

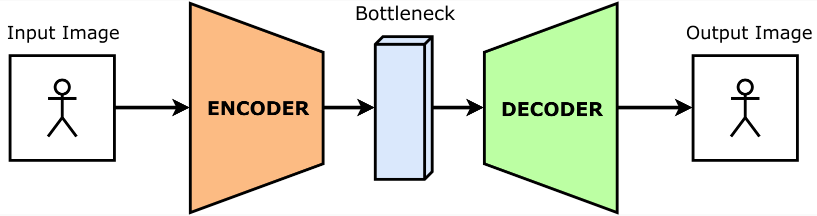
Basic perceptron Scheme:



$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Neural Networks Overview: Auto-Encoder

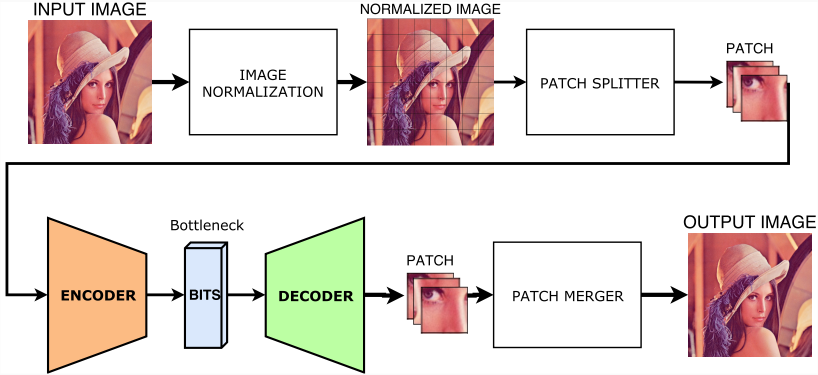
Auto-Encoder Basic Scheme:



[2] Depicted as images but any type of data can be used as input

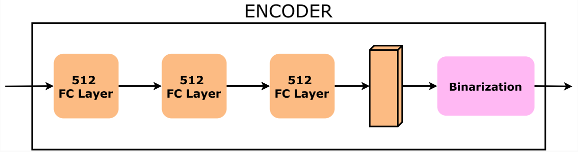
Architectures and Schematics

Architectures and Schematics: General Scheme

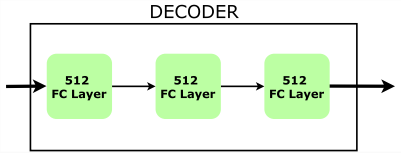


Architectures and Schematics: Straight-Forward FC

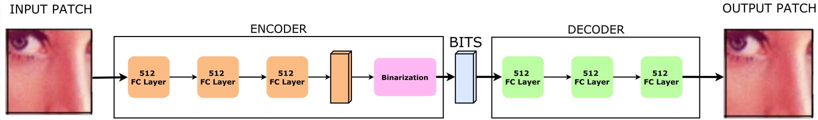
Encoder:



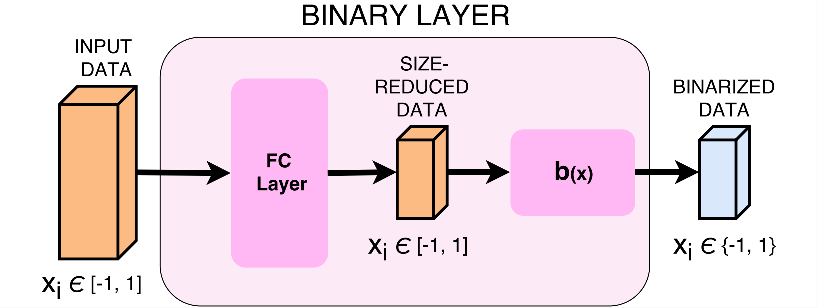
Decoder:



Straight-Forward Scheme:



Architectures and Schematics: Binary Layer



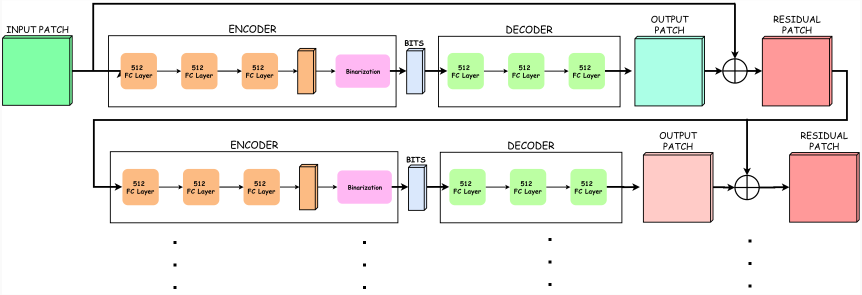
$$b(x) = x + \epsilon \ \in \{-1, 1\}$$

$$\epsilon \sim \begin{cases} +1 - x & \text{with probability } \frac{1+x}{2}, \\ -x - 1 & \text{with probability } \frac{1-x}{2}. \end{cases}$$

[3] Method proposed by: Raiko, Tapani, et al. "Techniques for Learning Binary Stochastic Feedforward Neural Networks." [1406.2989] ICLR 2015, 9 Apr. 2015, arxiv.org/abs/1406.2989

Architectures and Schematics: Residual FC

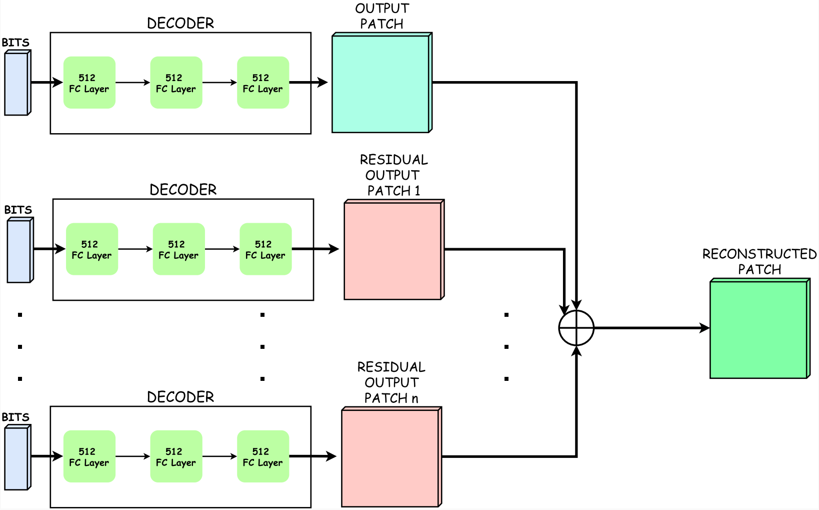
Encoding Scheme:



[4] Using tanh non-linearities in all layers

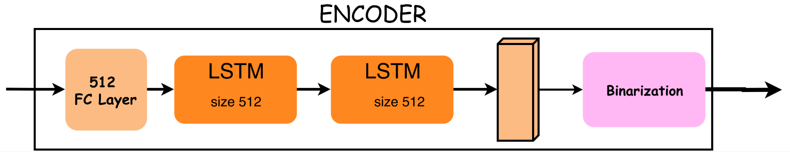
Architectures and Schematics: Residual FC

Decoding Scheme:

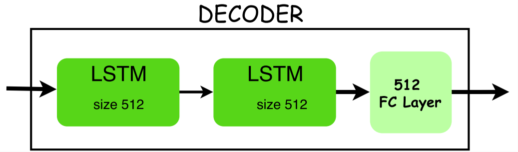


Architectures and Schematics: LSTM Based Scheme

Encoder Block:

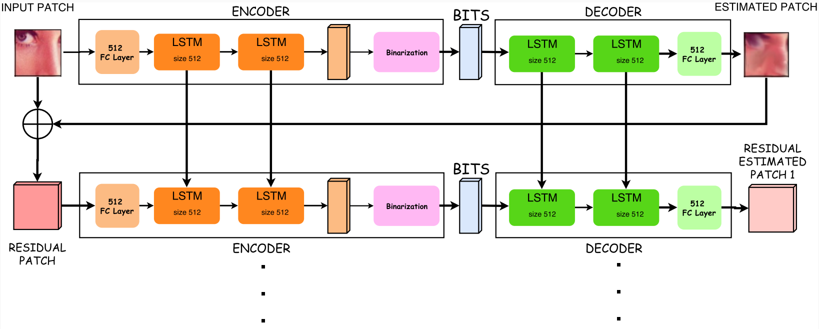


Decoder Block:



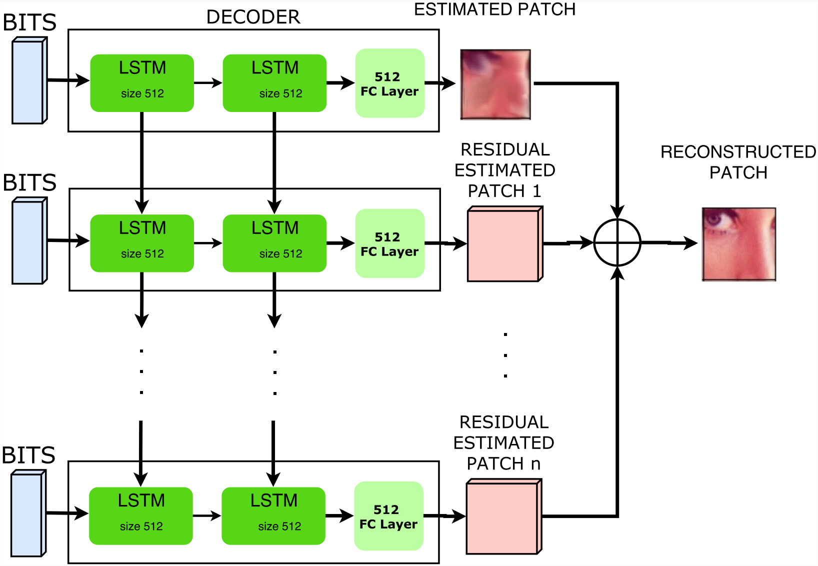
Architectures and Schematics: LSTM Based Scheme

Encoder:



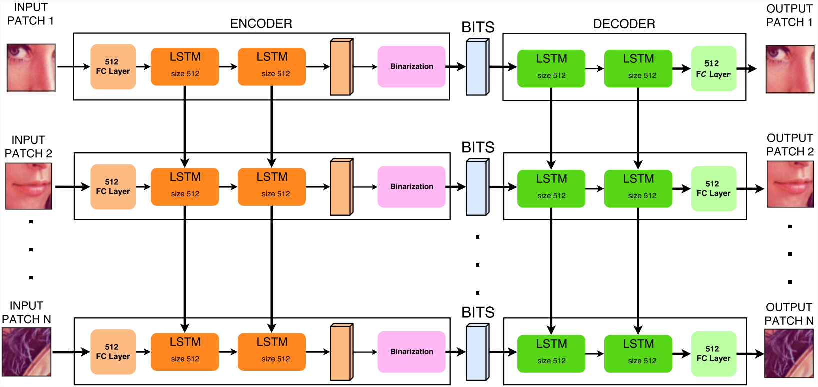
[5] Using tanh non-linearities in all layers

Decoder:


BITS
DECODER
ESTIMATED PATCH
LSTM
size 512
LSTM
size 512
512
FC Layer
BITS
RESIDUAL
ESTIMATED
PATCH 1
RECONSTRUCTED
PATCH
LSTM
size 512
LSTM
size 512
512
FC Layer
BITS
RESIDUAL
ESTIMATED
PATCH n
LSTM
size 512
LSTM
size 512
512
FC Layer

Architectures and Schematics: Our LSTM Proposed Scheme

Decoder:



Results

Results: Resources Used

- Framework:
- Computation:
- Dataset: *CIFAR10* and *CIFAR100*

Results: Results Overview

1. Straight-Forward Fully-Connected
2. Resiudal Fully-Connected
   - 16 bpp (2048 B)
   - 4 bpp (512 B)
   - 1 bpp (128 B)
   - 0.5 bpp (64 B)
3. Convolutional and Residual Convolutional
4. Our LSTM

Results: Parameters to Adjust

- Encoder-Decoder scheme
- Patch size
- Encoded size
- Number of passes
- Training parameters

Results: Little Quizz

What is in the picture?

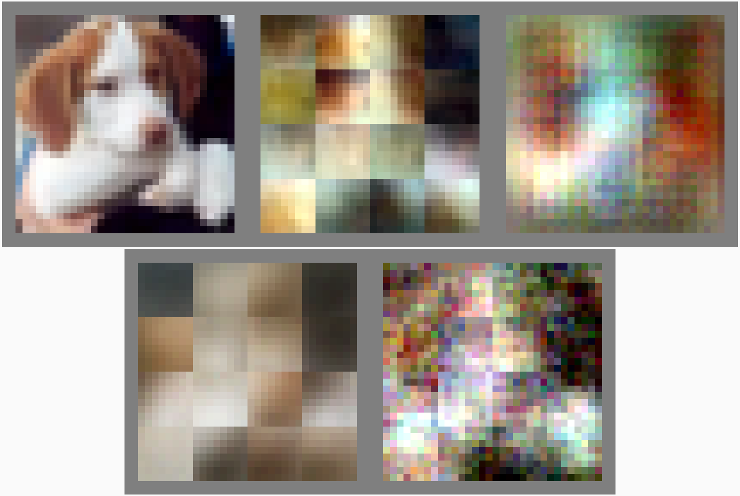
Results: Little Quizz

What is in the picture?

[6] 4bpp (512B) FC Residual decoded (left) and original (right) 32x32 images

Results: Poor-Quality Results Comparison

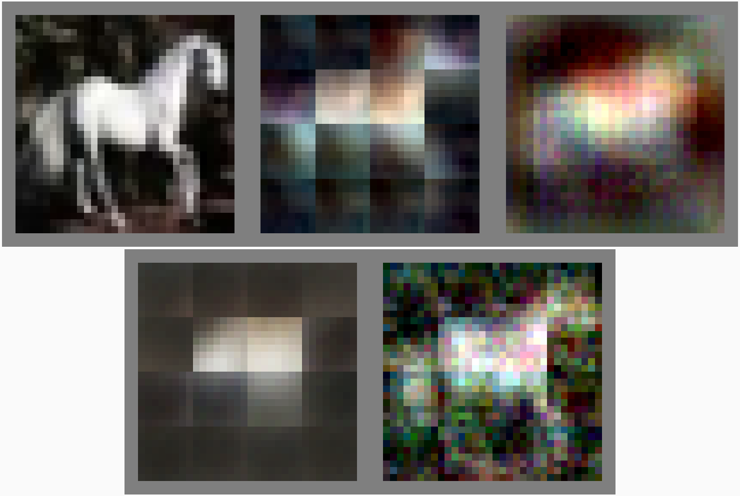
Results comparison using 0.5 bpp:



[7] From top-left to bottom-right:
Original, FC, Convolutional Residual, Our LSTM, FC Resiudal

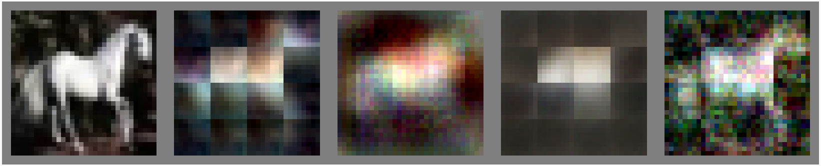
Results: Poor-Quality Results Comparison

Results comparison using 0.5 bpp:



[8] From top-left to bottom-right:
Original, FC, Convolutional Residual, Our LSTM, FC Resiudal

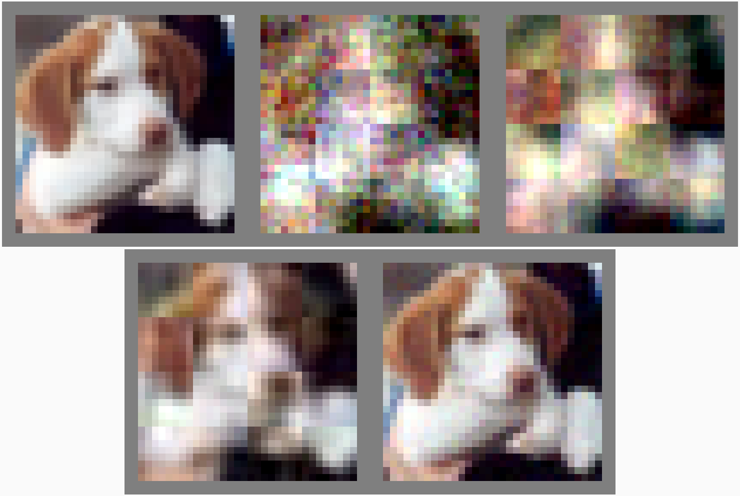
Results: Poor-Quality Results Comparison



	FC	Convolutional Residual	Our LSTM	FC Residual
SSIM	0.7844	0.6932	0.6520	0.8191
MSE	77.942	79.370	81.64	77.527

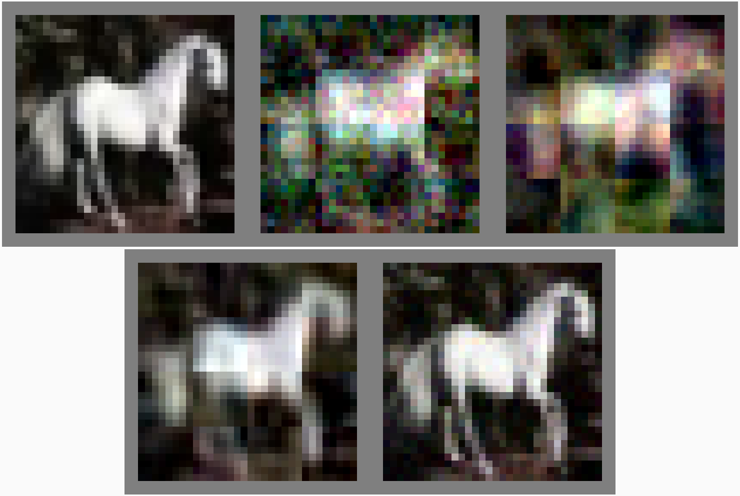
[9] Scores shown are the average value over a pre-defined set of 20 random pictures
Pictures from right to left: Original, FC, Convolutional Residual, Our LSTM and FC Residual

Results: FC Residual Analysis



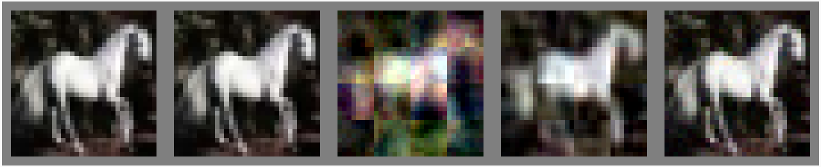
[10] From top-left to bottom-right: Original, 0.5 bpp, 1 bpp, 4 bpp 16 bpp

Results: FC Residual Analysis



[11] From top-left to bottom-right: Original, 0.5 bpp, 1 bpp, 4 bpp 16 bpp

Results: FC Residual Analysis



	JPEG	FC Residual		
BPP	1.20	1	4	16
SSIM	0.967	0.8654	0.9478	0.9951
MSE	30.06	74.351	64.952	20.451

[12] Scores shown are the average value over a pre-defined set of 20 random pictures
Pictures from right to left: Original, JPEG, FC Resiudual using 1, 4 and 16 bpp

Conclusions

Conclusions

- Low variable bitrate coding acomplish
- NN can be applied to image coding
- JPEG solid coding scheme and hard to be outperformed
- Results obtained are fair for an accademic project
- Further improves are needed

VRIC: Variable Rate Image Compression with Recurrent Neural Networks

Coding of Audiovisual Contents

Miquel Oller Oliveras & Alvaro Scherk Fontanals

Barcelona, December 18, 2017

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

telecom BCN
Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona