ICIP 2019 TP.PG.7

Image Pre-Transformation for Recognition-Aware Image Compression

Satoshi Suzuki, Motohiro Takagi, Kazuya Hayase, Takayuki Onishi, Atsushi Shimizu
NTT Media Intelligence Laboratories, NTT Corporation, JAPAN

NTT

Introduction

Images with high compression ratios degrade deep neural network (DNN) recognition accuracy

- ✓ DNNs automatically obtain a feature extraction mechanism and selectively **extract high frequency features** of images (e.g. Gabor like edge)
- ✓ Lossy image compression **discards the high frequency information** and often brings coding artifacts (e.g. block-shaped distortions, mosquito noises)

Our Goal : Prevent the DNN recognition accuracy degradation due to lossy compression

Basic Ideas

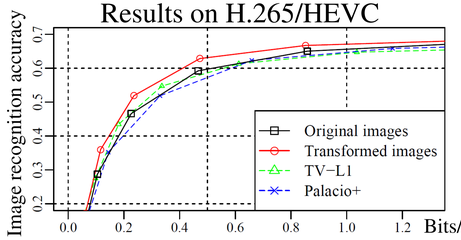
Increasing spatial correlation for reduce bitrates while maintaining accuracy using image pre-transformation

- ✓ Our method pre-transforms images before lossy compression
- ✓ Related work [Palacio+, CVPR18]
  - Their model pre-transform in images that gave them higher accuracy than the original one
  - The encoder-decoder (ED) model is learnt with the backpropagated loss of DNNs

Results

- ✓ **Higher accuracy** than some benchmarks in highly compressed situation
- ✓ **Good BD-Rate(Accuracy) gain** on various encoders

Results on H.265/HEVC



	Palacio+	TV-L1 [Le Guen, IPOL14]	Ours
JPEG	+3.7%	+0.4%	**-11.9%**
JPEG2000	+15.9%	+1.3%	**-17.8%**
HEVC	+9.5%	-3.9%	**-21.5%**

BD-Rate(Accuracy) on each encoder

	w/o TV	w/ TV
w/o bypass	+9.5%	-14.7%
w/ bypass	+4.2%	**-21.5% (Ours)**

Ablation results of our methods on HEVC

Experimental Conditions
Task: ImageNet classification problem
Network: VGG-16 [Simonyan & Zisserman, ICLR15]
QPs selection for BD-Rate: 4points under 1.0bpp
Hyper-parameter: $\lambda = 1.0$

Conclusion

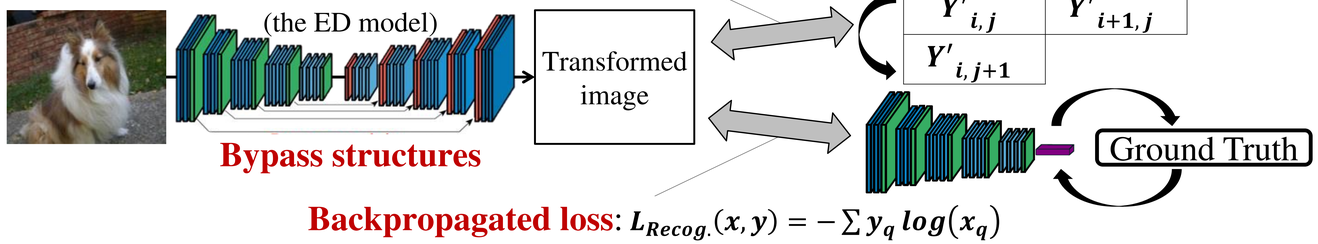
Our image pre-transformation model can prevent the DNN accuracy degradation due to compression on various encoders

- Our method can apply for JPEG and JPEG2000 directly, in contrast previous works [Choi & Bajic, ICASSP18] [Galteri+, ICPR18]
- Future works: further improve the loss functions and investigate our method's impact on recognition performance that humans can achieve

Our Method

- ✓ Proposed model is learnt with total variation loss and backpropagated loss
  $Loss = L_{Recog.}(x, y) + \lambda \cdot L_{TV}(Y')$ * λ: Hyper-parameter
  - Total variation (TV) loss has the effect of increasing the spatial correlation
  - Because it is **not differentiable** that directly calculate bitrate, we focus on the **spatial correlation** of images
- ✓ Our model is the ED model with bypass structures
  - Bypass structures can prevent degradation problem of DNN [He+, CVPR16] and make it possible to utilize local image features

TV loss: $L_{TV}(Y') = \sum \left(\left| Y'_{i+1,j} - Y'_{i,j} \right|^2 + \left| Y'_{i,j+1} - Y'_{i,j} \right|^2 \right)$



Backpropagated loss: $L_{Recog.}(x, y) = -\sum y_q \, log(x_q)$