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-transforms in images that gave them higher accuracy than the original one
  - The encoder-decoder (ED) model is learnt with the backpropagated loss of DNNs

Our Method

- Proposed model is learnt with total variation loss and backpropagated loss
  \[ \text{Loss} = L_{\text{Recog.}}(x,y) + \lambda \cdot L_{TV}(Y') \]
  - 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

Results

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

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Palacio+ [Le Guen, IPOL14]</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPEG</td>
<td>+3.7%</td>
<td>-11.9%</td>
</tr>
<tr>
<td>JPEG2000</td>
<td>+15.9%</td>
<td>-17.8%</td>
</tr>
<tr>
<td>HEVC</td>
<td>+9.5%</td>
<td>-21.5%</td>
</tr>
</tbody>
</table>

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

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

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