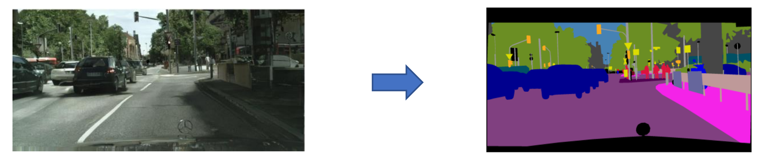
DSSLIC: Deep Semantic Segmentation-based Layered Image Compression

Mohammad Akbari, Jie Liang
School of Engineering Science, Simon Fraser University, Canada
akbari@sfu.ca, jiel@sfu.ca

Jingning Han
Google Inc.
jingning@google.com

Introduction

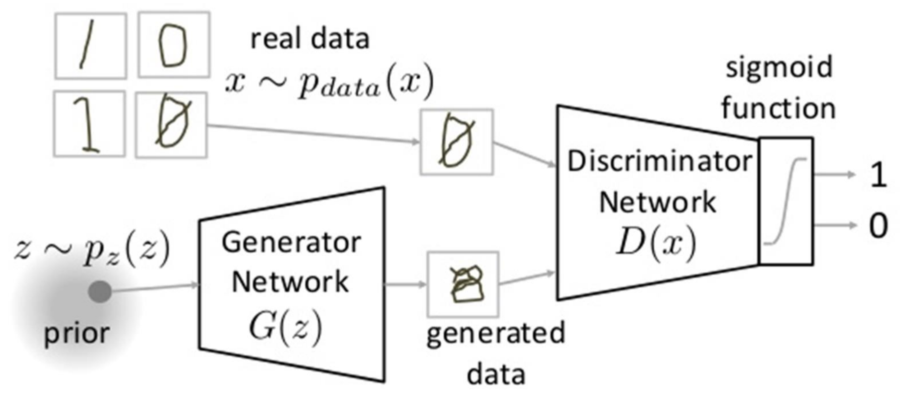
- **Deep learning in Computer Vision**
  - Image classification, face recognition
  - Semantic Segmentation (image to map) [1]



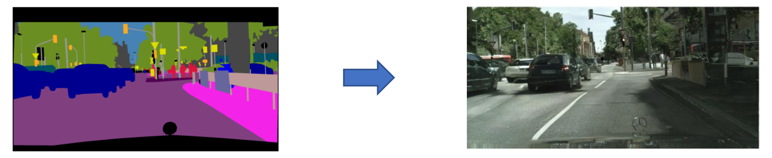
- **Generative Adversarial Networks (GANs)**
  - Competition of **D**iscriminator and **G**enerator

$$\min_G \max_D V(D,G)$$

$$V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))]$$

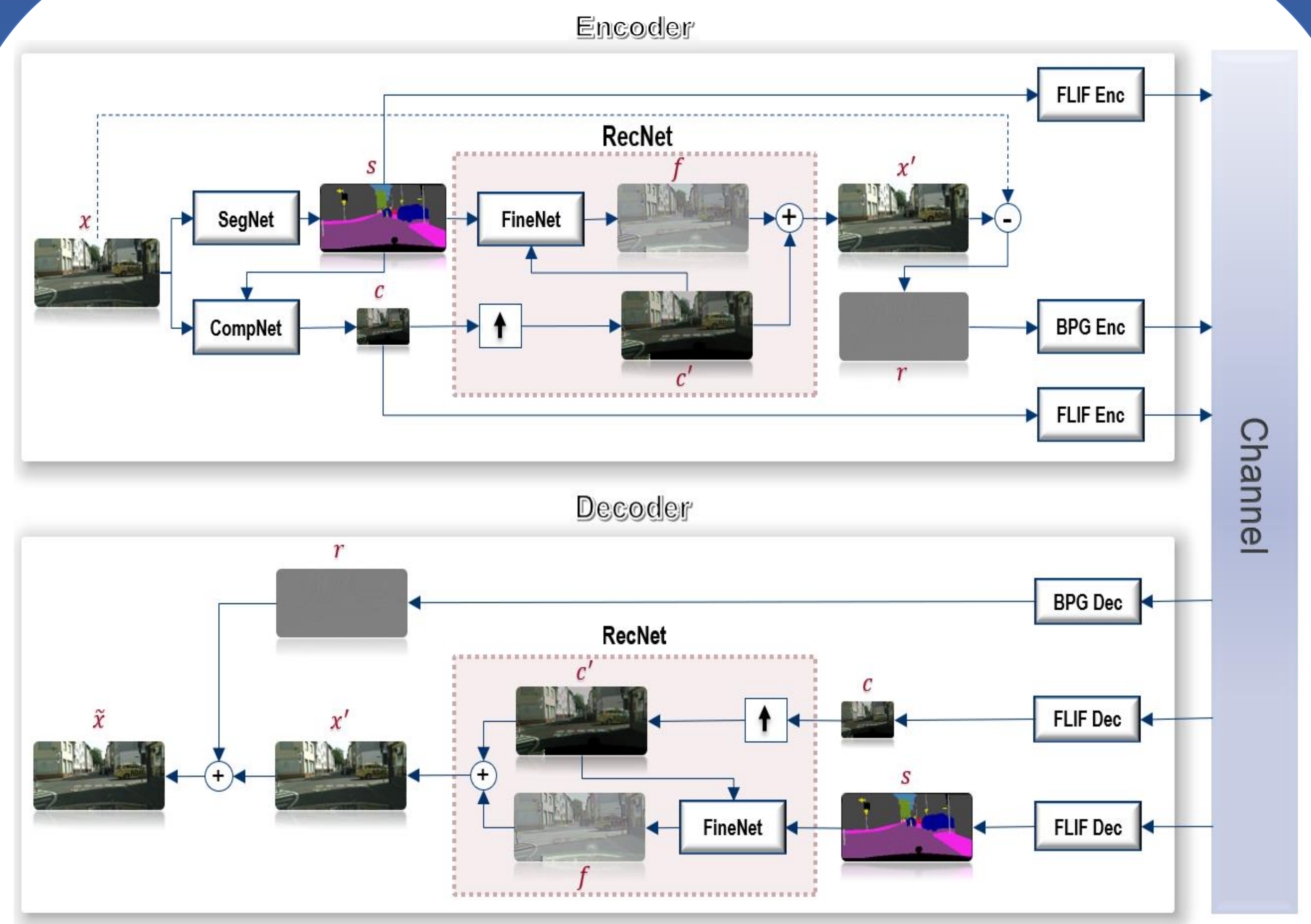


- GAN-based high-quality **map to image** [1]



- **Learning-based Image Compression (IC)** [2-5]
  - Exploit image features using neural networks
  - Learn analysis and synthesis transforms
  - Better results than standard codecs with hand-crafted components (e.g., JPEG & BPG)
- **Previous works:**
  - GAN-based autoencoder for IC [2]
  - Soft-to-hard vector quantization approach [3]
  - GDN-based analysis/synthesis transform [4]
  - Semantic map-based IC for low bit rates [5]

Proposed Approach



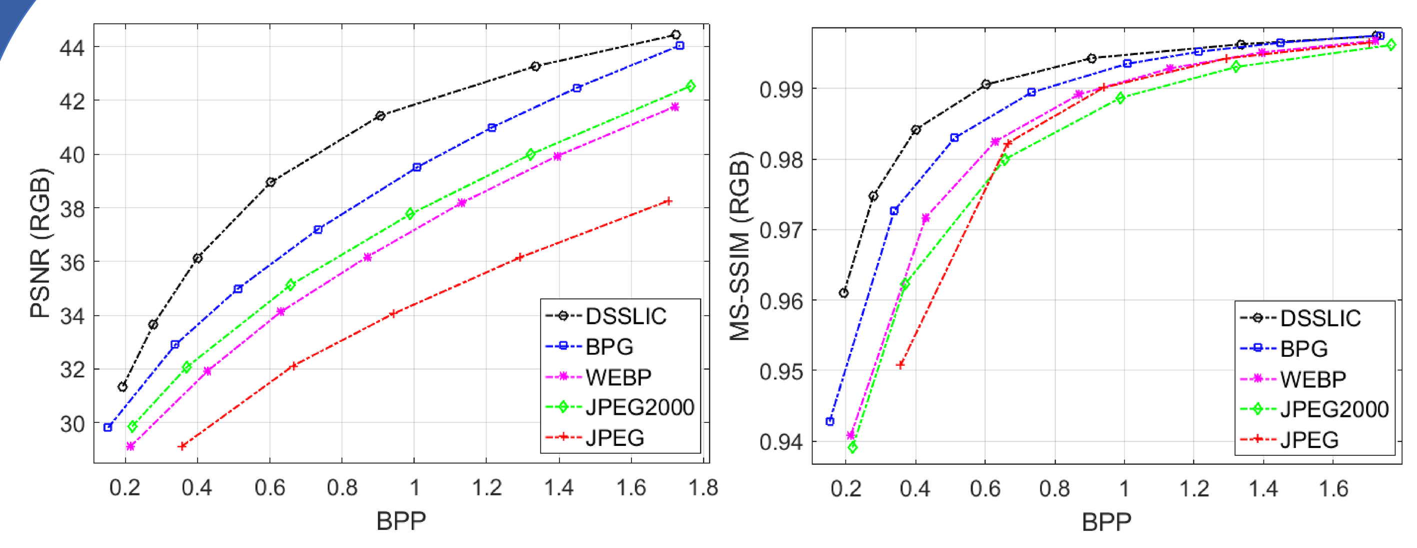
- 3 encoded layers: compact image **c**, semantic map **s**, residual **r**
- 3 deep networks: **SegNet**, **CompNet**, and **FineNet** (Generator)
- Using **GANs** with multi-scale discriminator
- **Objective Functions**:
  - Pixel-wise losses: $\mathcal{L}_1 = 2\lambda\|x - x'\|_1$. $\mathcal{L}_{SSIM} = -I(x,\tilde{x}).C(x,\tilde{x}).S(x,\tilde{x})$,
  - Perceptual losses: $\mathcal{L}_{DIS} = \lambda \sum_{d=1,2}\sum_{i=1}^{n}\frac{1}{N_i}\|D_d^{(i)}(s,c',x) - D_d^{(i)}(s,c',x')\|_1$,

    $\mathcal{L}_{VGG} = \lambda\sum_{j=1}^{m}\frac{1}{M_j}\|V^{(j)}(x) - V^{(j)}(x')\|_1$,
  - Adversarial training:

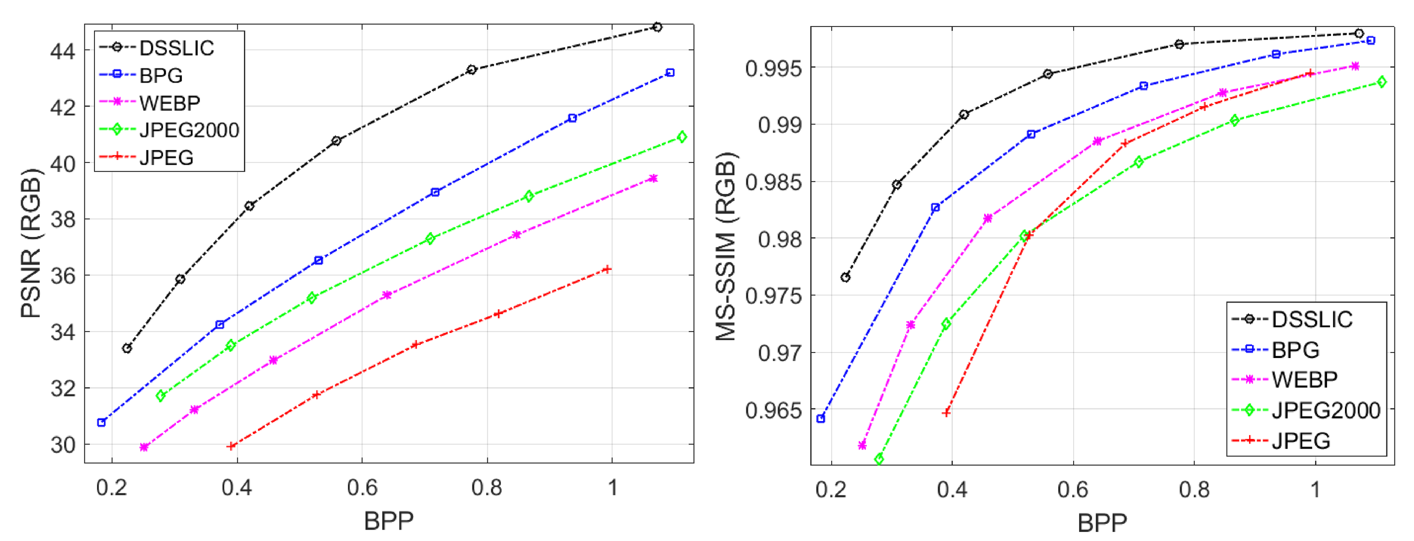
$$\left.\begin{aligned}\mathcal{L}_D &= -\sum_{d=1,2}(\log D_d(s,c',x) + \log(1-D_d(s,c',x'))),\\ \mathcal{L}_G &= -\sum_{d=1,2}\log D_d(s,c',x') + \mathcal{L}_1 + \mathcal{L}_{SSIM} + \mathcal{L}_{DIS} + \mathcal{L}_{VGG}.\end{aligned}\right\} \quad \mathcal{L} = \mathcal{L}_D + \mathcal{L}_G.$$

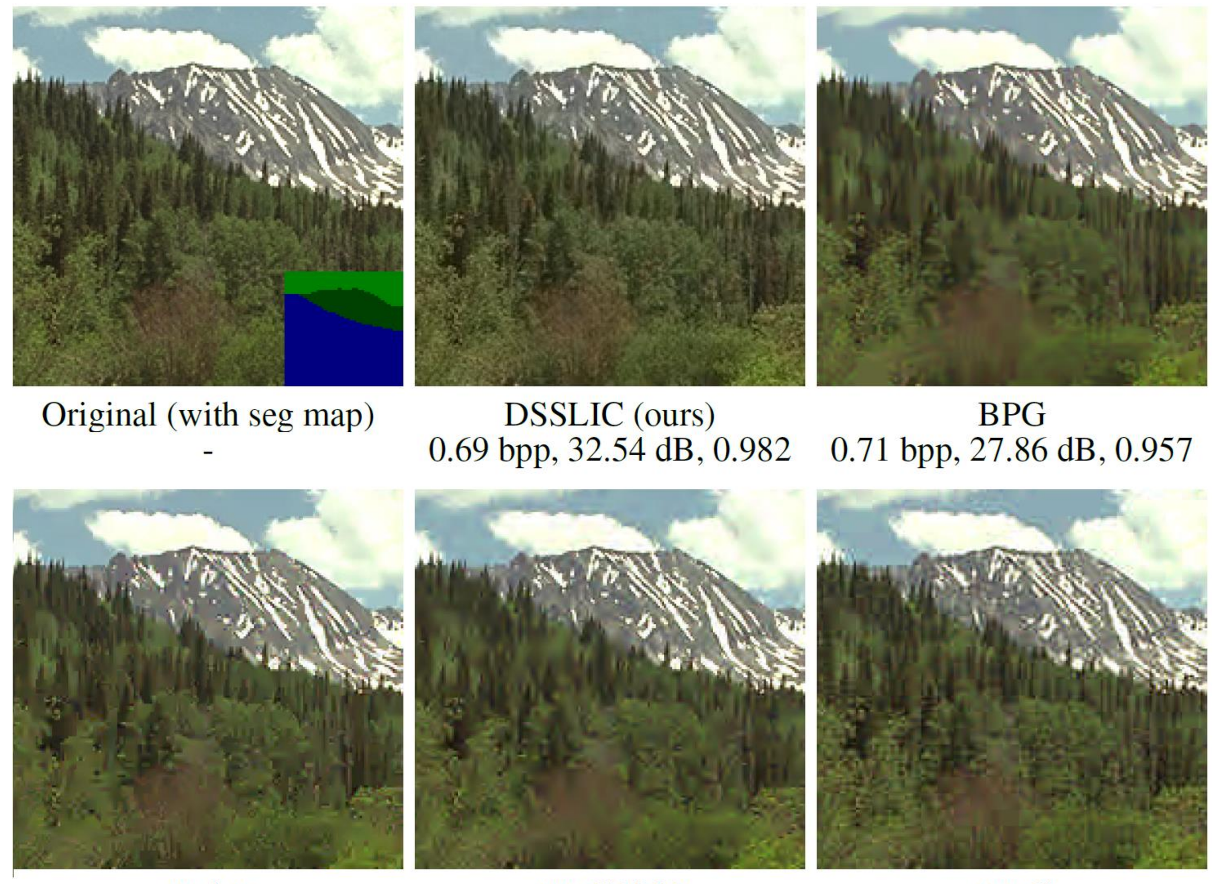
Experimental Results

Comparison results on Kodak

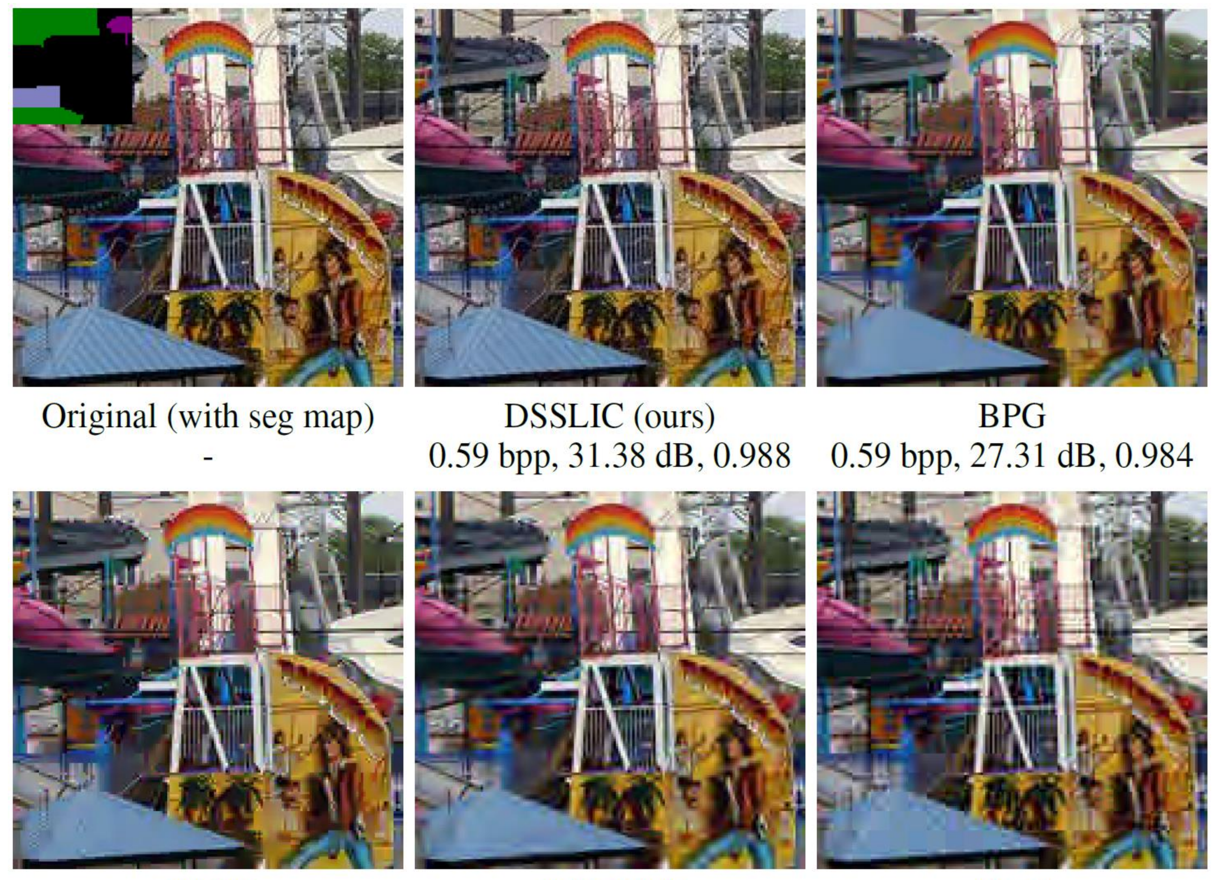


Comparison results on ADE20K





Original (with seg map) - | DSSLIC (ours) 0.69 bpp, 32.54 dB, 0.982 | BPG 0.71 bpp, 27.86 dB, 0.957
WebP 0.71 bpp, 26.01 dB, 0.952 | JPEG2000 0.71 bpp, 26.71 dB, 0.942 | JPEG 0.72 bpp, 24.77 dB, 0.958



Original (with seg map) - | DSSLIC (ours) 0.59 bpp, 31.38 dB, 0.988 | BPG 0.59 bpp, 27.31 dB, 0.984
WebP 0.60 bpp, 25.43 dB, 0.979 | JPEG2000 0.60 bpp, 25.12 dB, 0.972 | JPEG 0.61 bpp, 23.63 dB, 0.973

Conclusion

- Deep learning-based image compression method proposed using semantic segmentation maps.
- Adversarial training using GAN + perceptual losses.
- Better PSNR and MS-SSIM than all standard codecs
- **Future works**:
  - YUV-based image coding
  - Object-based adaptive image compression
  - Deep scalable image compression

References

1. Wang et al. "High-resolution image synthesis and semantic manipulation with conditional GANs." CVPR 2018.
2. Oren et al. "Real-time adaptive image compression." ICML 2017.
3. Eirikur et al. "Soft-to-hard vector quantization for end-to-end learning compressible representations." NIPS 2017.
4. Minnen et al. "Joint autoregressive and hierarchical priors for learned image compression." NIPS 2018.
5. Eirikur et al. "Extreme learned image compression with GANs." CVPR Workshops 2018.