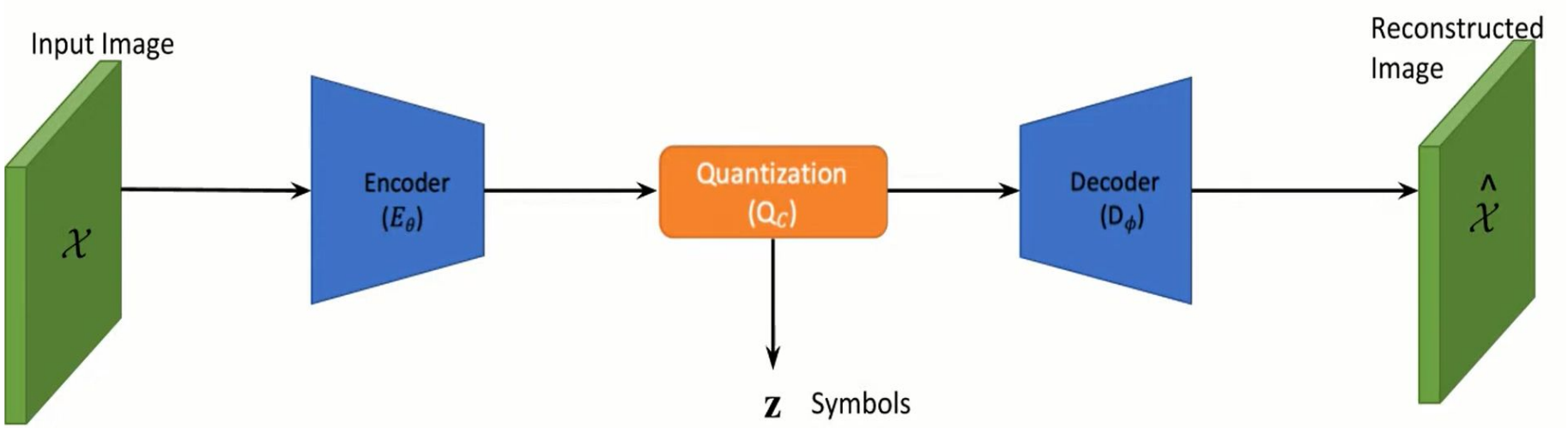
Saliency Driven Perceptual Image Compression

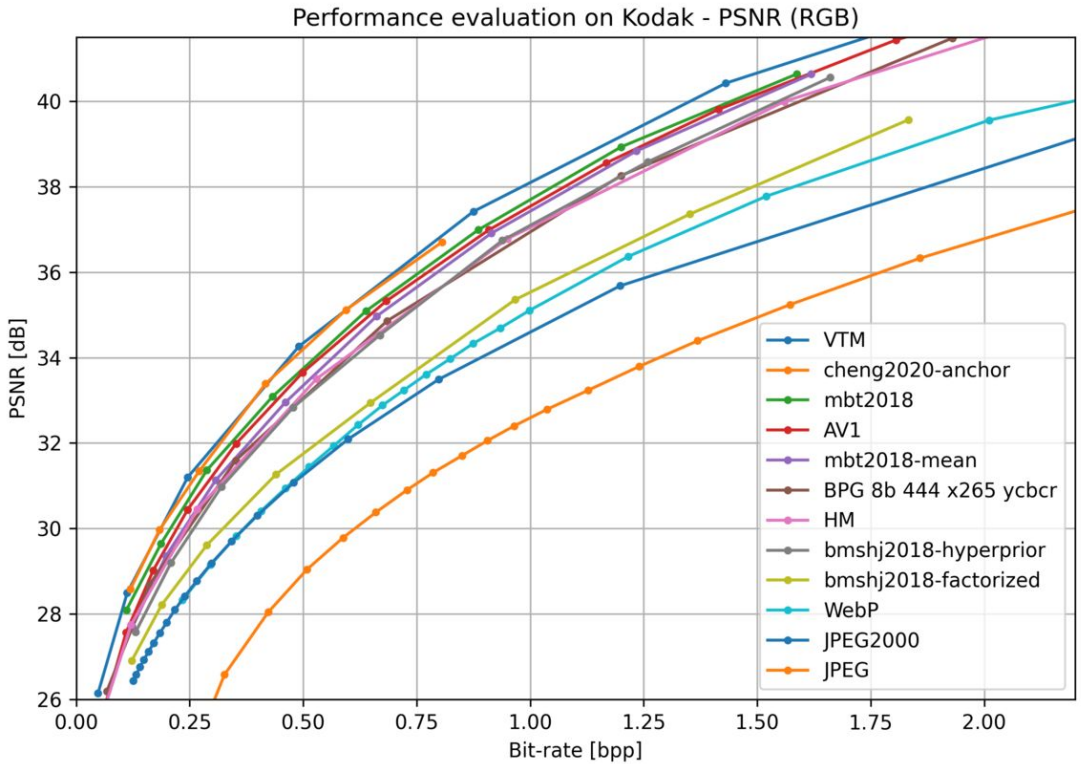
Yash Patel , Srikar Appalaraju , R. Manmatha Amazon Web Services, Palo Alto Visual Recognition Group, Czech Technical University in Prague

Trénovanie neurónových sietí na kompresiu obrázkov



aky je ciel trenovania modelu?

Rate-distortion krivka



Problém zhodnotenia kvality po rekonštrukcii

Štrukturálna podobnosť (SSIM)

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

Stredná kvadratická chyba (MSE) a špičkový pomer signálu k šumu (PSNR)

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\|I(i,j) - K(i,j)\|^2$$

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) = 20 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I}{\sqrt{\mathrm{MSE}}}\right)$$

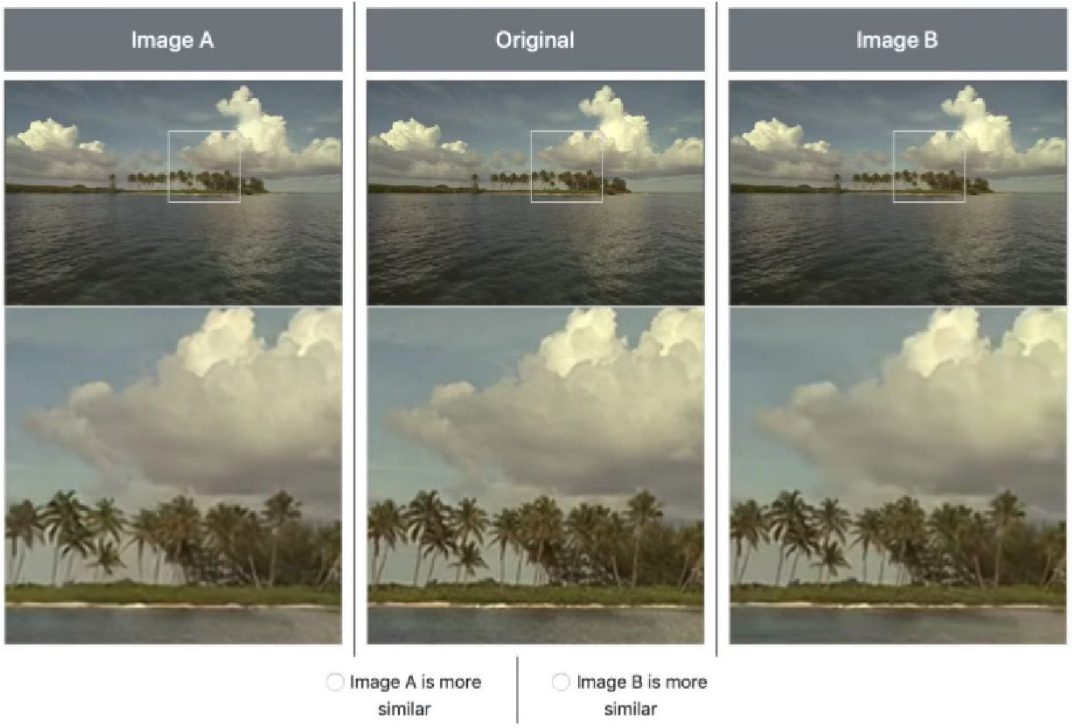
trenovanim optimalizujeme hodnoty tychto metrik

Higher MS-SSIM

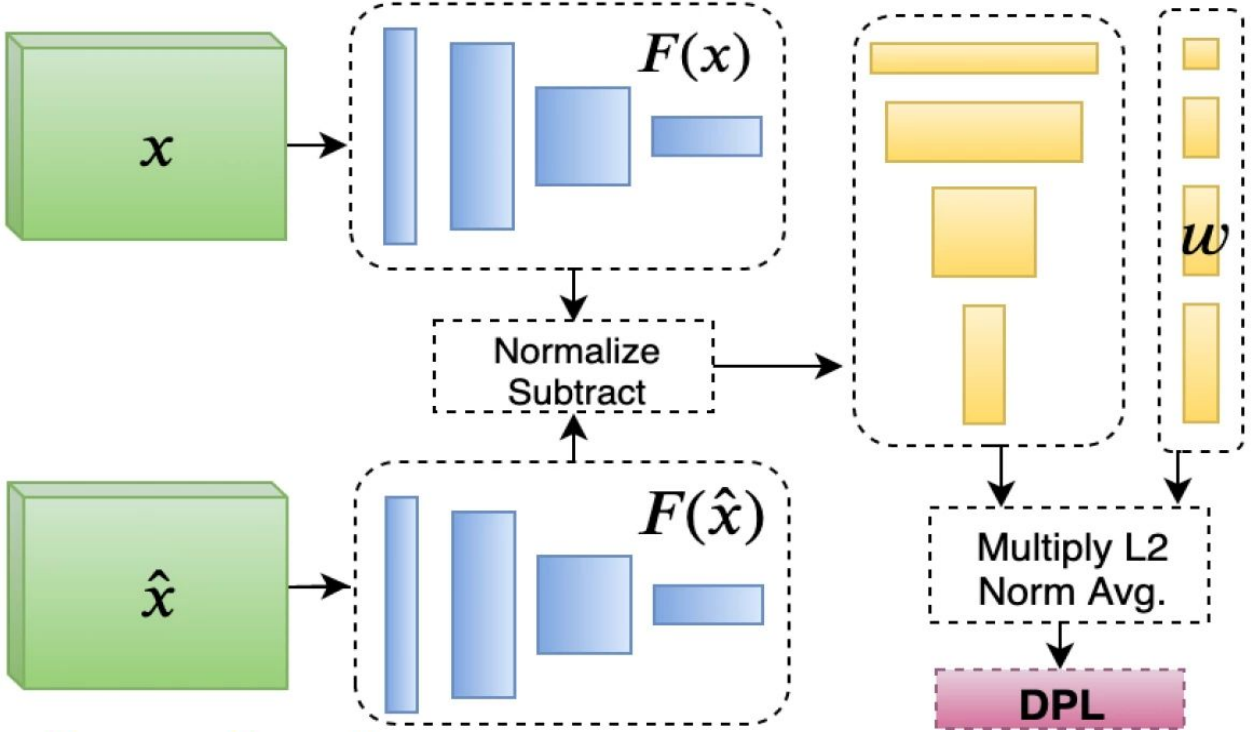
Original | Mentzer et al. | Ballé et al. | BPG | JPEG-2000

Perceptual similarity dataset

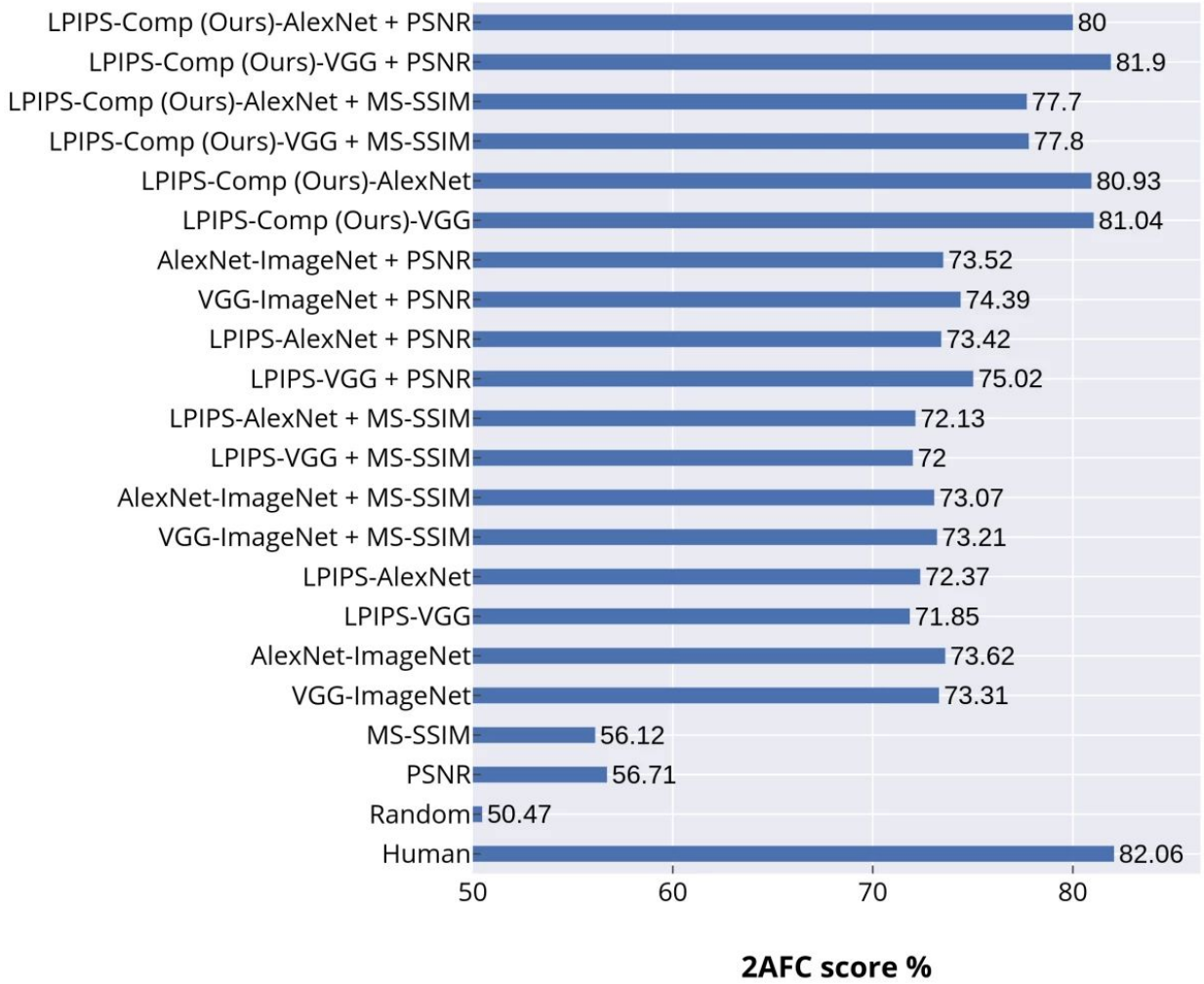
two alternatives forced choices (2AFC)



Aproximácia ľudského vnímania kvality - compression specific, learnt perceptual similarity metric (deep perceptual loss)



$$DPL(\mathbf{x}, \hat{\mathbf{x}}) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} ||\mathbf{w_l} \odot (\mathbf{z^l_{\hat{x},h,w}} - \mathbf{z^l_{x,h,w}})||_2^2 \quad (1)$$


LPIPS-Comp (Ours)-AlexNet + PSNR
80
LPIPS-Comp (Ours)-VGG + PSNR
81.9
LPIPS-Comp (Ours)-AlexNet + MS-SSIM
77.7
LPIPS-Comp (Ours)-VGG + MS-SSIM
77.8
LPIPS-Comp (Ours)-AlexNet
80.93
LPIPS-Comp (Ours)-VGG
81.04
AlexNet-ImageNet + PSNR
73.52
VGG-ImageNet + PSNR
74.39
LPIPS-AlexNet + PSNR
73.42
LPIPS-VGG + PSNR
75.02
LPIPS-AlexNet + MS-SSIM
72.13
LPIPS-VGG + MS-SSIM
72
AlexNet-ImageNet + MS-SSIM
73.07
VGG-ImageNet + MS-SSIM
73.21
LPIPS-AlexNet
72.37
LPIPS-VGG
71.85
AlexNet-ImageNet
73.62
VGG-ImageNet
73.31
MS-SSIM
56.12
PSNR
56.71
Random
50.47
Human
82.06
50
60
70
80
2AFC score %