



Semantic Perceptual Image Compression

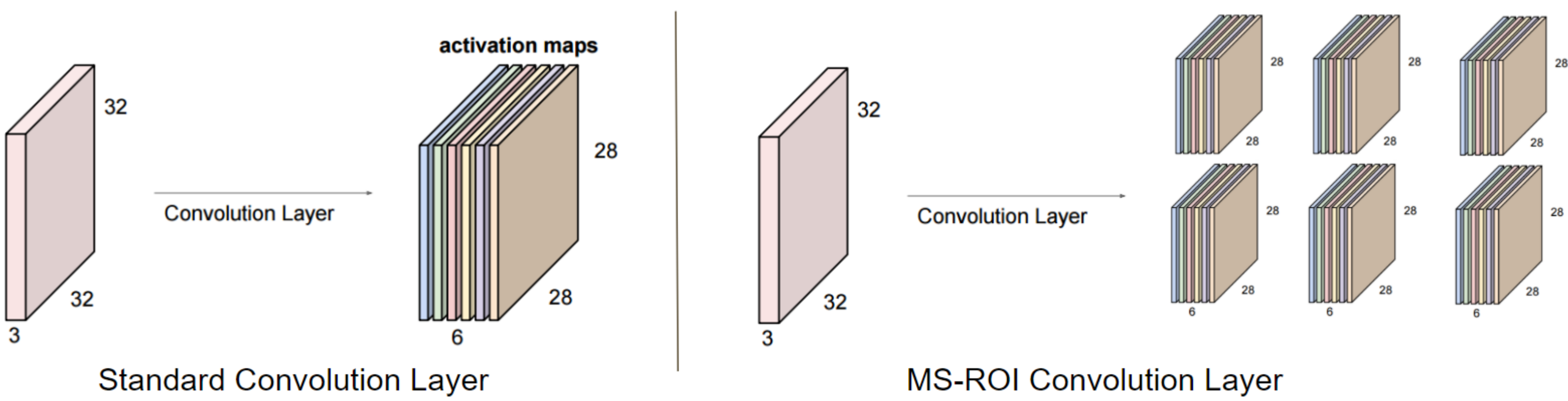
Adi Prakash, N. Moran, S. Garber, A. DiLillo & J. Storer

✉ aprakash@brandeis.edu, iamaaditya.github.io

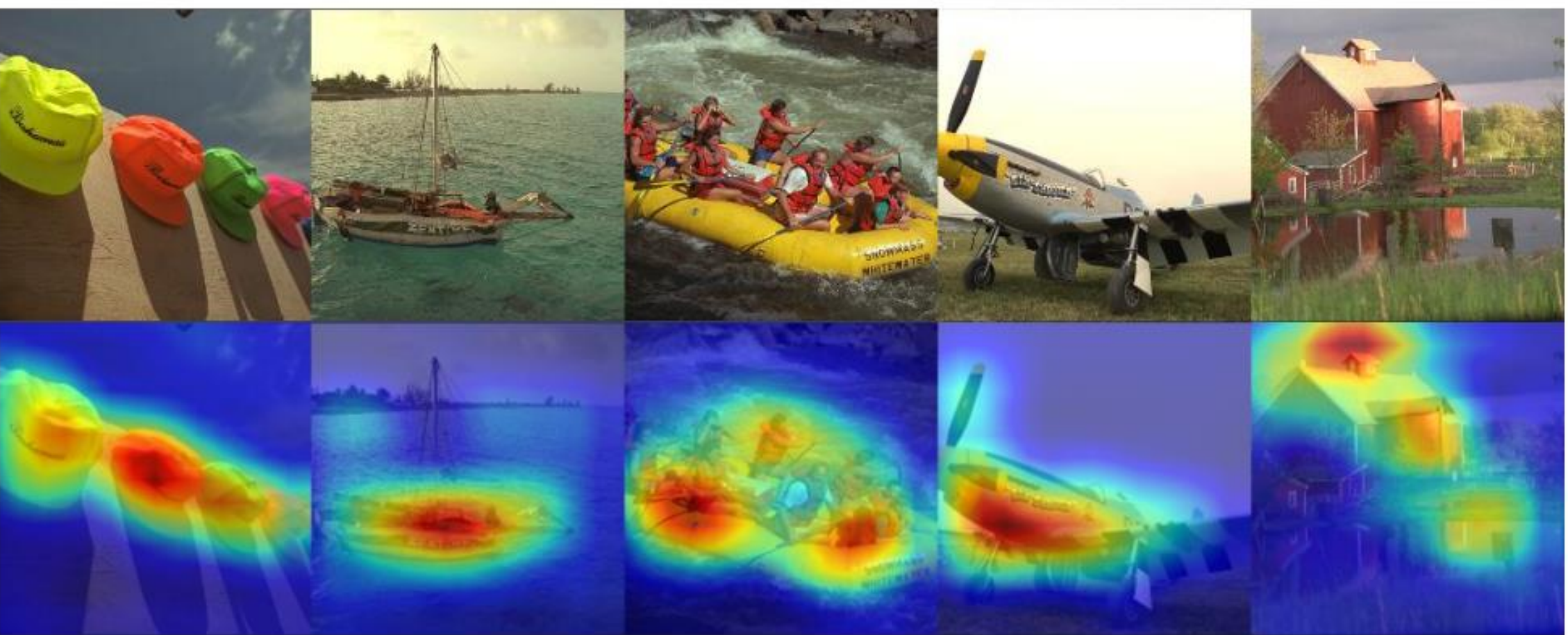
🔗 github.com/iamaaditya/image-compression-cnn

IEEE DCC
SLC - 2018

Multi-Structure Region of Interest



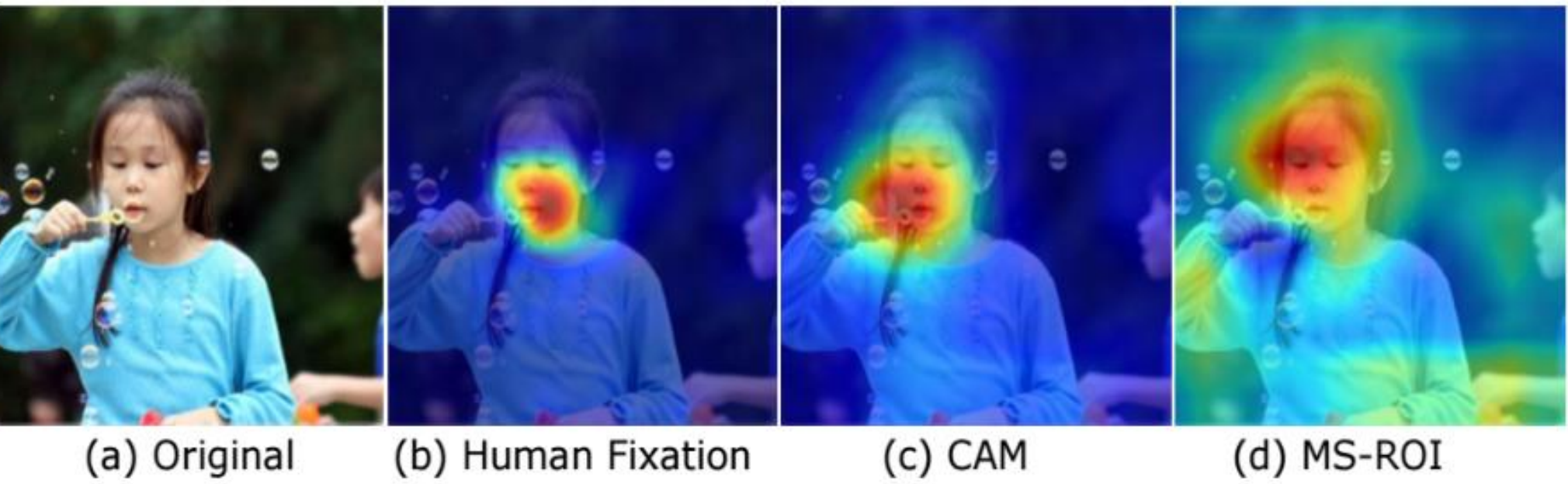
MSROI Examples



Results

	PSNR-S	PSNR	PSNR-HVS	PSNR-HVSM	SSIM	MS-SSIM	VIFP
Kodak PhotoCD [24 images]							
Std JPEG	33.91	34.70	34.92	42.19	0.969	0.991	0.626
Our model	39.16	34.82	35.05	42.33	0.969	0.991	0.629
MIT Saliency Benchmark [Outdoor Man-made + Natural, 200 images]							
Std JPEG	36.9	31.84	35.91	45.37	0.893	0.982	0.521
Our model	40.8	32.16	36.32	45.62	0.917	0.990	0.529
Re-sized images of a very large image, see fig: 4 [20 images]							
Std JPEG	35.4	27.46	33.12	43.26	0.912	0.988	0.494
Our model	39.6	28.67	34.63	44.89	0.915	0.991	0.522

MSROI - Details



- Z_l^c denotes threshold which signifies 'presence' of a class
- \widehat{M} denotes Multi-structure map generated using MSROI. Compare this with CAM map (M)
- It is sum over all classes with total activations Z_l^c beyond some threshold.
- For training use sigmoid instead of softmax to prevent losing information about 'other objects'

CAM

$$M_c(x, y) = \sum_{d \in \mathbf{D}} w_d^c f_d(x, y)$$

where w_d^c is learned for every class c and for layer ' d '

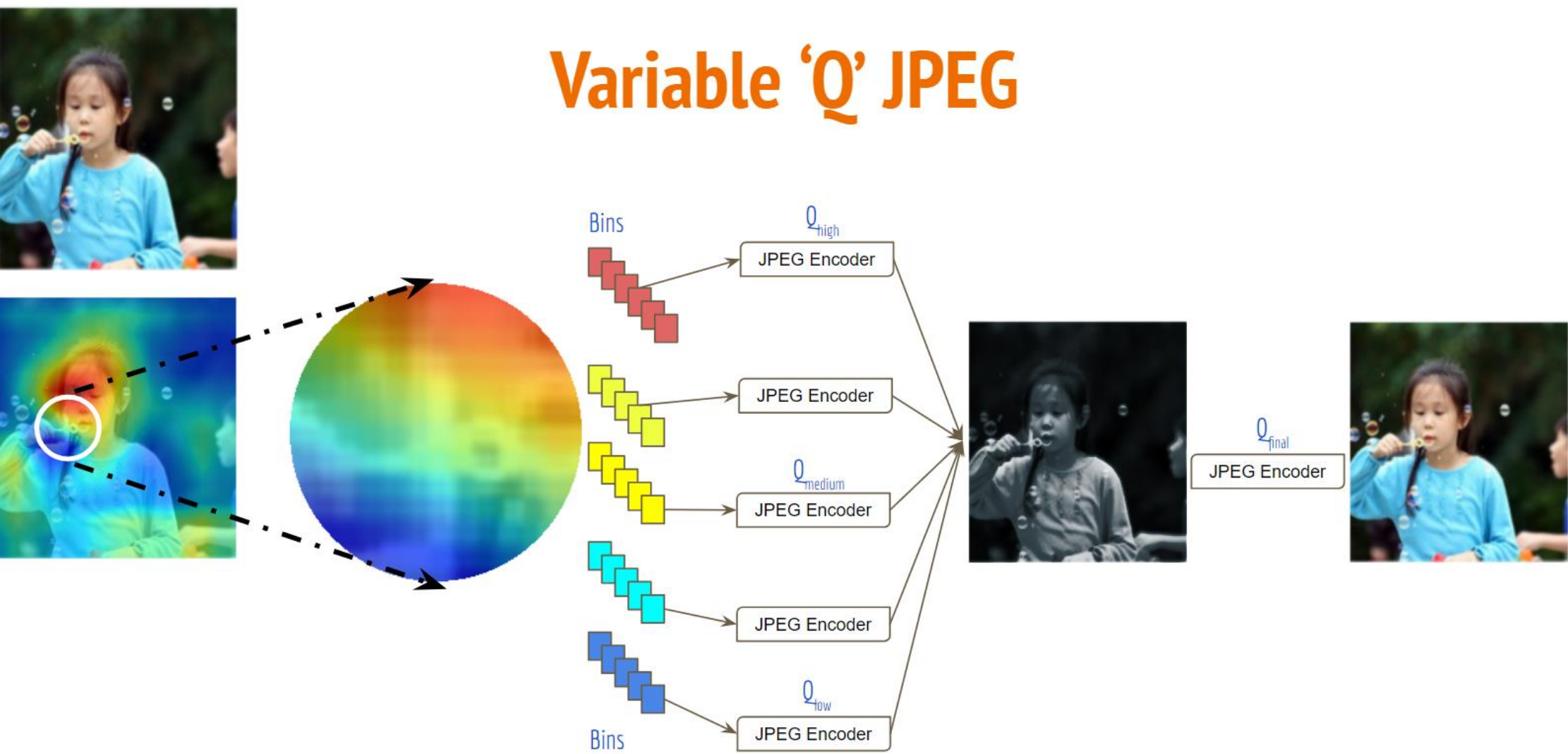
$$P(c) = \frac{\exp(\sum_{xy} M_c(x, y))}{\sum_c \exp(\sum_{xy} M_c(x, y))}$$

MSROI Map

$$Z_l^c = \sum_{d \in \mathbf{D}} \sum_{x, y} f_d^c(x, y)$$

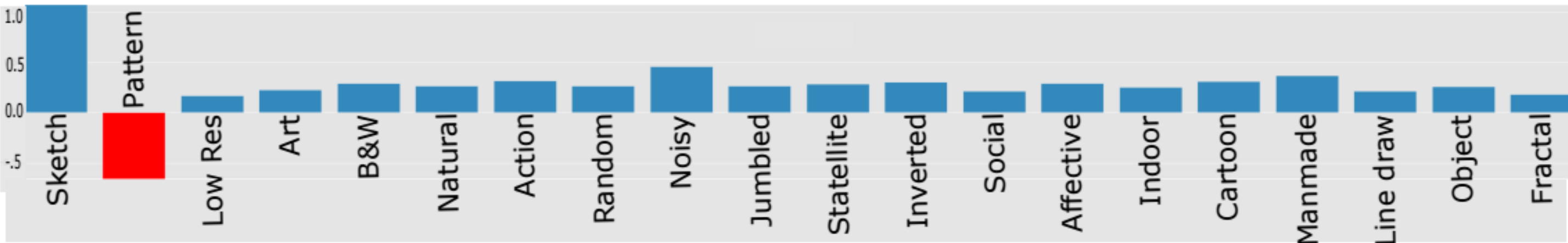
$$\widehat{M}(x, y) = \sum_{c \in \mathbf{C}} \begin{cases} \sum_d f_d^c(x, y), & \text{if } Z_l^c > T \\ 0 & \text{otherwise} \end{cases}$$

$$P(c) = \frac{1}{1 + \exp(Z_l^c)}$$



PSNR-S is the PSNR of the 'salient' regions as identified by MSROI

Comparison of different categories



Comparison of different resolutions

