

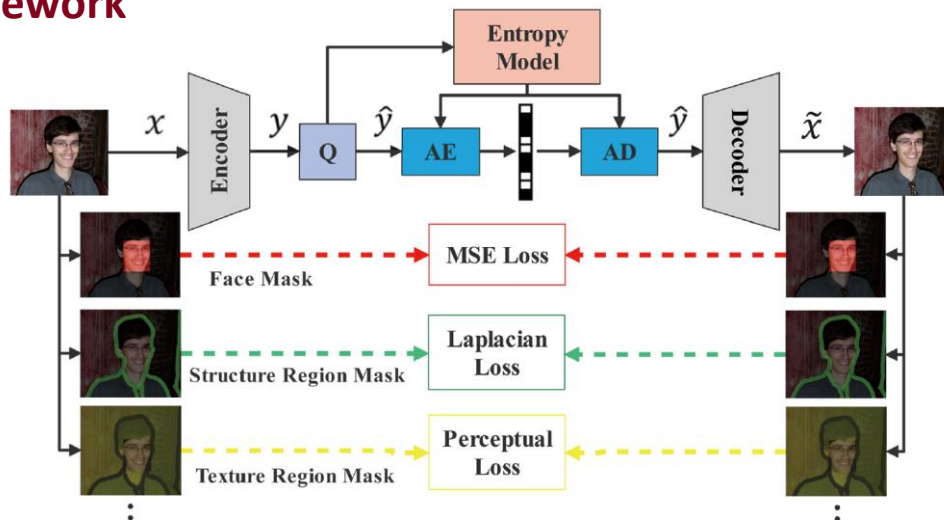
## Main Idea

Image content plays an important role in human perception, and human eyes have different sensitivities to different content. However, the influence of image content has been largely ignored in learning-based image compression. Different regions of the images have different properties, so the training strategy should also be different.

## Contribution

We propose a content-oriented E2E image compression scheme, which can be used in most of existing methods and no extra information is needed in the encoding and decoding stage.

## Framework



$$\mathcal{L}_{total} = \eta R(\hat{y}) + \epsilon \mathcal{L}_{stru} + \mathcal{L}_{per} + \gamma \mathcal{L}_{sface}$$

Laplacian loss is used for structure region to recover the shape edges. Perceptual loss including MAE loss, LPIPS loss, and GAN loss is adopted in the texture region. Besides, MSE is used for the facial regions for an accurate reconstruction. And the priority of the masks is different, specifically, from high to low: the facial mask, the structure mask and the texture mask. Note that all the masks are only used in the training phase, which means the method won't bring extra parameters.

## Ablation



## Visual Comparison



## Statistics Result

