Abstract:

Image restoration algorithms are typically evaluated by some distortion measure (e.g. PSNR, SSIM, IFC, VIF) or by human opinion scores that quantify perceived perceptual quality. In this work, we prove mathematically that distortion and perceptual quality are at odds with each other. Specifically, we study the optimal probability for correctly discriminating the outputs of an image restoration algorithm from real images. We show that as the mean distortion decreases, this probability must increase (indicating worse perceptual quality). As opposed to the common belief, this result holds true for any distortion measure, and is not only a problem of the PSNR or SSIM criteria. However, as we show experimentally, for some measures it is less severe (e.g. distance between VGG features). We also show that generative-adversarial-nets (GANs) provide a principled way to approach the perception-distortion bound. This constitutes theoretical support to their observed success in low-level vision tasks. Based on our analysis, we propose a new methodology for evaluating image restoration methods, and use it to perform an extensive comparison between recent super-resolution algorithms. Our study reveals which methods are currently closest to the theoretical perception-distortion bound.  
* Joint work with Yochai Blau.