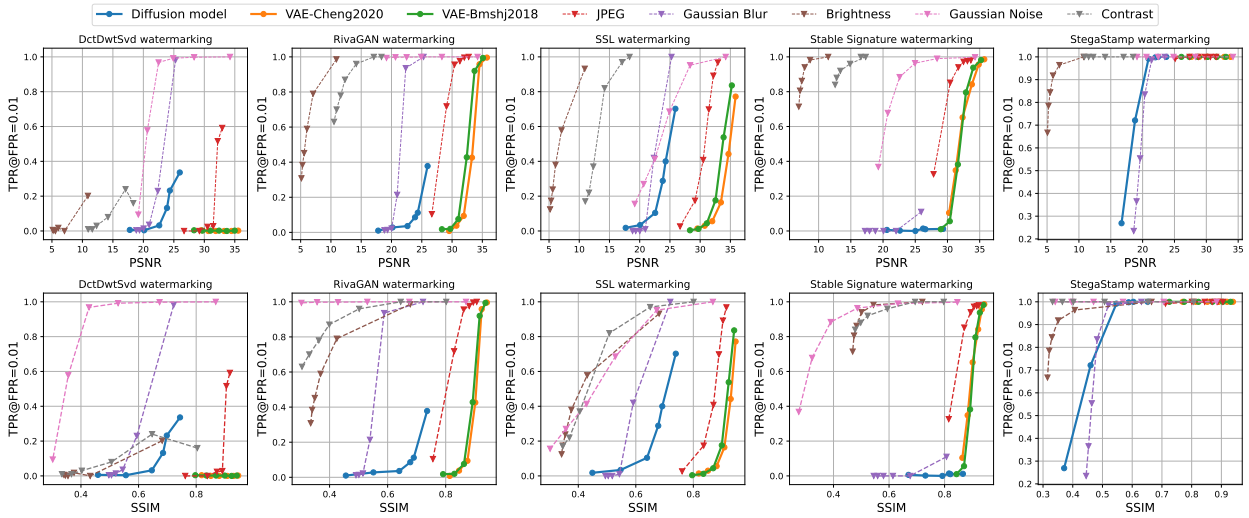

Figure 1: Quality-detectability tradeoff for five watermark schemes under eight attack methods. The x-axis shows image quality metrics (SSIM and PSNR, higher values indicate better quality), while the y-axis represents the detection metric True Positive Rate at 1% False Positive Rate (TPR@FPR=0.01, lower values are better for attackers). The strongest attacker should appear in the lower right corner of these plots.

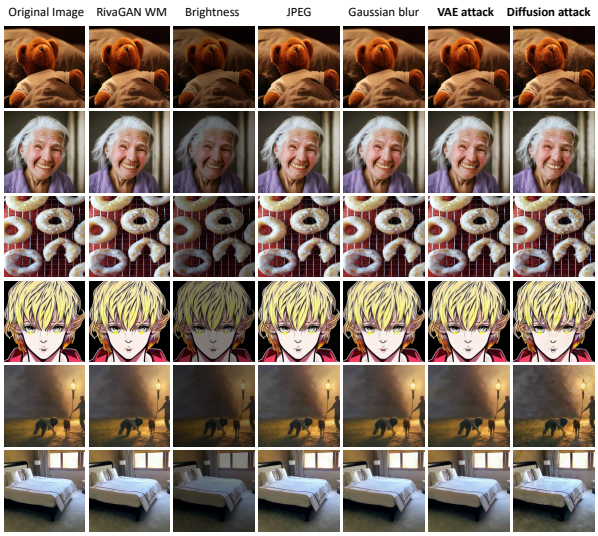

Figure 2: Examples of different attacks against RivaGAN watermark.

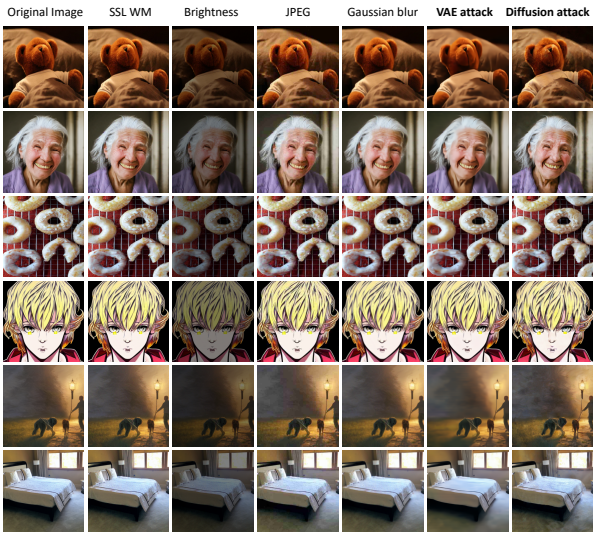

Figure 3: Examples of different attacks SSL watermark.

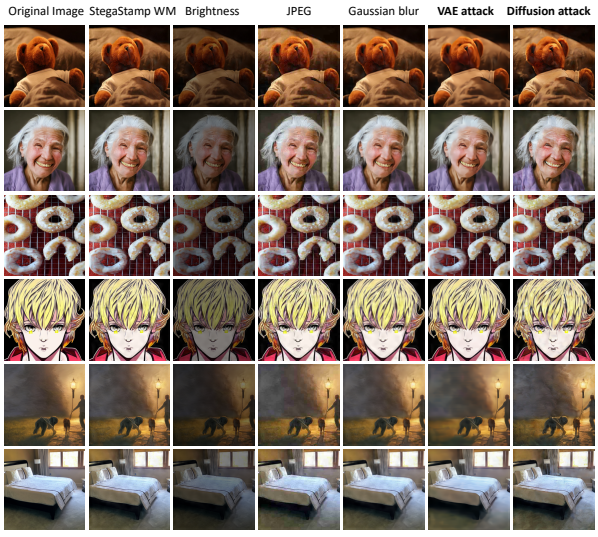

Figure 4: Examples of different attacks against StegaStamp watermark.

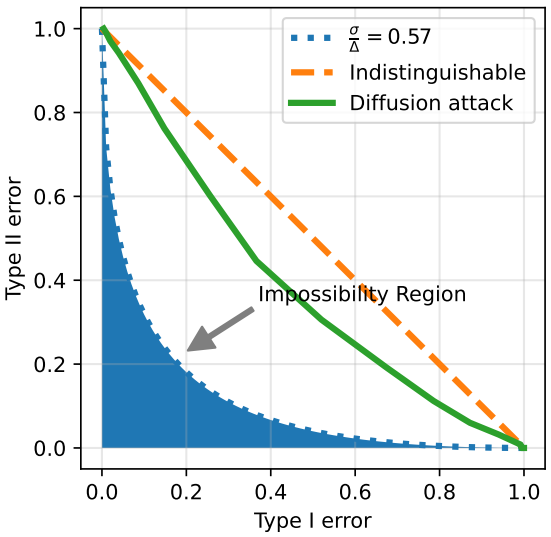

Figure 5: Theoretical and empirical trade-off functions for RivaGAN watermark.