Author : Mateo ZOUGHEBI

Title: «Watermarked image generative models. Case study: GAN models»

Short Abstract:

AI-generated content (images, videos, audio, text, etc.) is expanding rapidly, driven by recent scientific breakthroughs and their integration into user-friendly applications. The resulting transformations affect multiple dimensions of daily life, ranging from the emergence of new business verticals to concerns about sovereignty, disinformation, and environmental impact. In this context, our work has three related yet complementary targets: (i) the watermarking of image generative models, (ii) the watermarking of the images thus generated, and (iii) the compliance of these solutions with multimedia compression standards. To this end, our study encompasses three main steps. First, a comprehensive testing procedure allowing for the objective benchmarking of state-of-the-art watermarking solutions has been designed and developed, thus identifying their main limitations, in terms of key-size, down-scaling with respect to the model size, and systematic visual impact in the generated images. Secondly, the study focuses on issues related to the deployment of high-quality generative AI models and advances an on/off image generation solution. This solution establishes synergies among state-of-the-art concepts to advance a new training scheme to enable the deployment of a GAN that can generate both vanilla (high-quality) samples and imperceptibly marked samples. The illustrations concern StyleGAN2-ADA trained on Celeb-A. The experimental results evaluate imperceptibility (in term of FID and SSIM) for both unmarked and marked samples, and robustness against compression neural-network–based attacks, and pixel-domain image modifications. Synergies with emerging ISO/IEC AWI 21617-3 Media asset watermarking standards are established. Our experimental results bring a new possibility for trustworthy AI deployment in edge–cloud settings, where storage capacity is a major challenge, by deploying on/off watermarked GEN-AI model.

Télécom SudParis • Site d'Évry : 9 rue Charles Fourier - 91011 Évry-Courcouronnes Cedex - France • Site de Palaiseau : 19 place Marguerite Perey - 91120 Palaiseau France • Tél. +33 (0)1 60 76 40 40 • Siret : 180 092 025 00055 • NDA 11910803791 • APE : 8542Z
www.telecom-sudparis.eu