Deep Learning-Based Imperceptible Image Watermarking with Residual Autoencoder and Perceptual VGG Loss

Dr. R. M. Krishna Sureddi
Associate Professor
Dept. of Information Technology
CBIT, Hyderabad, India
rmkrishnasureddi_it@cbit.ac.in

Rama Devi N
Assistant Professor
Dept. of Information Technology
CBIT, Hyderabad, India
nrd@cbit.ac.in

Pujitha Badrigari
Student
Dept. of Information Technology
CBIT, Hyderabad, India
bpoojitha32@gmail.com

Veena Ayyannagari
Student
Dept. of Information Technology
CBIT, Hyderabad, India
veenaayyannagari.az@gmail.com

Dasari Navyakrishna
Student
Dept. of Information Technology
CBIT, Hyderabad, India
navyakrishna010@gmail.com

***Abstract*—The increasing use of digital media to spread the message that privacy protection, information protection, preservation of image metadata,protection against AI-generated images , and copy rights should be promoted more strongly in this sector.We provide a blind image watermarking method using autoencoders architecture and Residual Block method, and a perceptual loss based on a VGG19 network. The binary watermark is embedded in the encoder of the network in the cover image, while the decoder can extract the watermark with high fidelity. The results demonstrated that a trade-off between imperceptibility and robustness was achieved as evaluated with PSNR, SSIM, and BER. We also examine new watermarking applications to identify AI-generated content and preserve image metadata In fact, this study is inspired by recent studies, such as the Res-Ception block architecture whilst focusing on the design of their Watermarking method to be lightweight, improving the performance, robustness and image quality of watermarked images. This paper provides an overview of the latest techniques that will enable the development of more complex and flexible media security watermarking models.**

***Index Terms*—Imperceptible watermarking, , Residual Blocks, Autoencoder, VGG19, Perceptual Loss, Deep Learning, PSNR, SSIM, BER ,Fidelity, Metadata Encryption, Colour watermarking, SVD, DCT, ND-Adam, DRM, CNN, DWT**

I. Introduction

A. Tradition Techniques

Frequency domain techniques insert watermarks by altering frequency co- efficients rather than spatial pixels . The typical transforms are Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), and Singular Value Decomposition (SVD).Graph-based watermarking models data as graphs, where each node repre- sents a pixel or feature, and edges capture relationships. These are especially useful for irregular data like 3D meshes, audio, or semantic regions.

Graph Fourier Transform (GFT) allows watermarking in the spectral domain of graphs. Graph Attention Networks (GATs) dynamically weight node features to better learn contextual importance. Such models are used in semantic segmentation, where watermarking needs to focus on meaningful regions Graph-based and attention-augmented watermarking techniques are a novel and developing field in digital watermarking research.

Watermarking of sensor-derived images (e.g., CMOS, thermal, SAR) has to cope with noise, low contrast, and hardware-specific distortions. Signal transforms and statistical models (e.g., LIST-DWT, PHPMs) facilitate the extraction of robust features in such environments. Low-light and hardware-embedded watermark detection utilizes custom CNNs such as cosine-convolutional networks. These are lightweight and tailored for embedded platforms Statistical and image sensor-based watermarking techniques are specifically designed for hardware-generated content like digital cameras, CMOS image sensors, thermal sensors, and remote sensing devices.

Encryption-based watermarking schemes add an extra layer of security by modifying the watermark data prior to embedding. Such manipulations render it computationally impossible for the adversary to extract or eliminate the watermark without the key. Typical techniques involve chaotic maps (such as logistic map), stream ciphers, hash functions, and Galois Field arithmetic. Chaotic systems are sensitive to the initial conditions and yield pseudo- random sequences that are best suited for watermark bit scrambling. Residue Number Systems (RNS) and orthogonal vector coding also enhance the resistance of watermark encoding.

Deep learning watermarking employs neural networks to insert and extract watermarks from digital content. These networks are trained end-to-end to learn the best representations for watermark data hiding and recovery. Typical architectures involve convolutional neural networks (CNNs), autoencoders, and adversarial networks (GANs). Recent methods incorporate

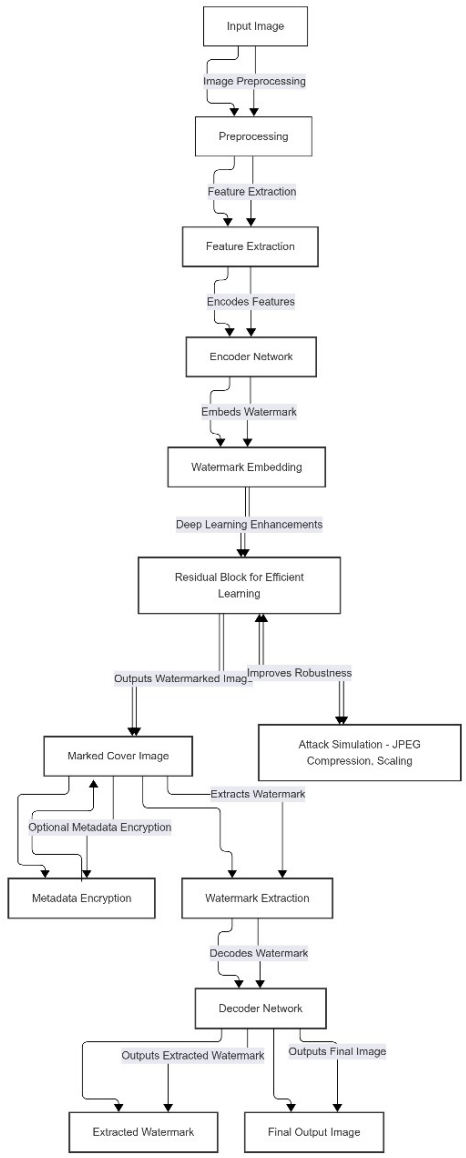
attention mecha- nisms and perceptual loss functions (e.g., VGG-based) to achieve a trade-off between imperceptibility and robustness. ND-Adam keeps the direction of gradients intact and normalizes the magnitudes of updates. Mini-batch training with real images and distorted images and JPEG simulation layers enhance robustness. Mini-batch methods are also employed to enhance generalization. Training is done on mini-batches comprising real and synthetically corrupted images rather than just clean images alone.

B. Exisiting Model

An effective and smart watermarking system based on convolutional autoencoders. This includes designing an encoder to insert a watermark into a cover image and a decoder to recover it correctly. – The aim is to make the watermarked (stego) image visually indistinguishable from the original image while enabling reliable watermark recovery. Maintainperceptibility and Robustness – The system must maintain the imperceptibility of the watermark, whiwhich means thats that the embedded watermark should not visibly alter the image. – Meanwhile, it should be robust, that is, it should be able to extract the watermark even when the image is subjected to usual distortions such as compression, noise, or minor editing. Loss functions such as Mean Squared Error and Binary Crossentropy assist in balancing both requirements during training. Performance evaluation using standard metrics, PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index measure) evaluate the visual quality of the stego image, while BER (bit error rate) evaluates the correctness of the extracted wa- termark. These parameters assist in verifying the practical usability and reliability of the system.

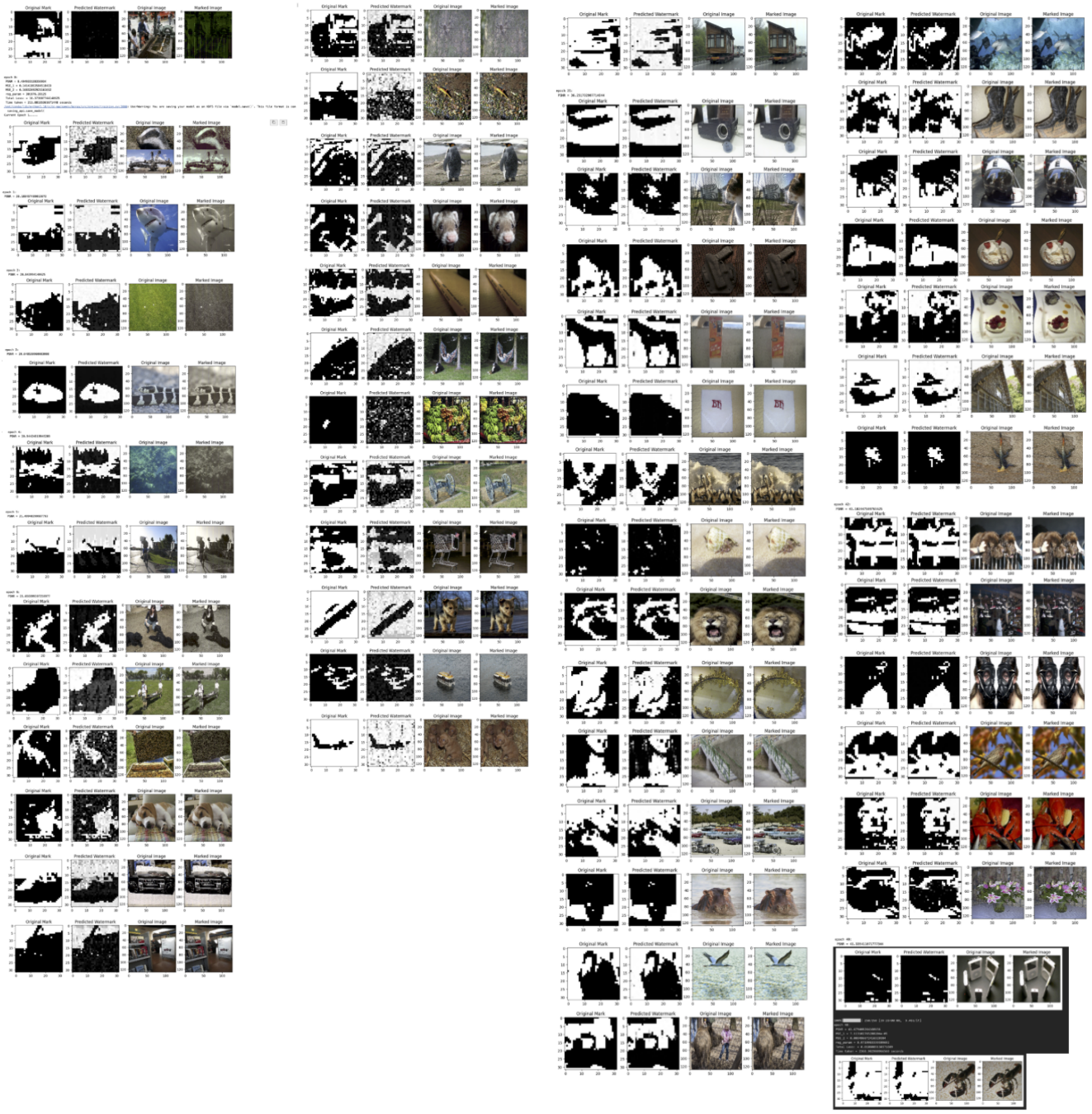
C. applications

Watermarking is an indispensable process for the real-time integrity, security, and reduced computation for IoT, autonomous vehicles, remote sensing and multimedia forensics. In IoT we could embed watermarks as a component of the real-time integrity on smart camera and drone data utilizing lightweight CNNs ensuring authentication and tamper resistance. SAR imaging will require robust strategies that can incorporate noise and signal-only variations. Medical imaging often implies a DICOM compliant watermark which may impose requirements enabling secure diagnostics.



(a) Deep Learning-based Image Watermarking Flowchart

(b) Results



(a) Results