Abstract

Reversible data hiding (RDH) techniques have garnered significant attention for their ability to embed data into digital media while ensuring the reversibility of the embedding process. This paper proposes a novel approach to RDH using Convolutional Neural Networks (CNNs), leveraging their capabilities in feature extraction and learning hierarchical representations. Unlike traditional methods that often rely on pixel-wise modifications, CNN-based RDH explores the feasibility of embedding data directly into feature maps or other intermediate representations extracted by deep convolutional networks.

This research aims to address several key challenges and opportunities in the field of CNN-based RDH:

1. **Embedding Capacity and Quality**: By harnessing the robust feature representation of CNNs, the proposed method seeks to enhance embedding capacity while minimizing perceptual distortion in the cover media.
2. **Adversarial Robustness**: Investigating methods to maintain the security and integrity of embedded data against adversarial attacks, ensuring robustness in real-world applications.
3. **Reversibility and Fidelity**: Developing efficient algorithms to achieve high-quality reconstruction of both the cover media and embedded data, ensuring negligible distortion after data extraction.

This study builds upon foundational research in reversible data hiding and deep learning, such as the works by Ma et al. (2018) on deep learning for steganography [1] and Fridrich et al. (2014) on reversible data hiding in images [2]. The proposed CNN-based approach aims to extend these principles by integrating state-of-the-art convolutional architectures with novel embedding strategies[3] & [4], thereby pushing the boundaries of reversible data hiding in digital media.

Through comprehensive experimental evaluations and comparisons with existing methods, this research aims to demonstrate the effectiveness and practicality of CNN-based RDH, paving the way for advanced applications in secure communication, content authentication, and multimedia forensics.

References:

[1] Ma, J., Li, W., Li, S., & Zhang, X. (2018). Deep steganography: End-to-end learning of content-adaptive hiding with deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[2] Fridrich, J., Goljan, M., & Du, R. (2014). Lossless data embedding for all image formats. IEEE Transactions on Image Processing, 23(3), 1481-1490.

[3] X. Yang and F. Huang, "New CNN-Based Predictor for Reversible Data Hiding," in *IEEE Signal Processing Letters*, vol. 29, pp. 2627-2631, 2022, doi: 10.1109/LSP.2022.3231193.

[4] R. Hu and S. Xiang, "CNN Prediction Based Reversible Data Hiding," in *IEEE Signal Processing Letters*, vol. 28, pp. 464-468, 2021, doi: 10.1109/LSP.2021.3059202.