Analyzing the effects of Alternative Watermarking Techniques on Deep Learning Computer Vision Tasks

Date: 11 Feb 2025

1. Project Summary

1.1. Team members.

(1) Sean Moulton, UMBC CMSC PhD student, seanmoulton@umbc.edu
(2) Anupreet Singh, UMBC CMSC Master's student, anuprel1@umbc.edu
(3) Omkar Kulkarni, UMBC CMSC PhD student, omkar.kulkarni@umbc.edu

1.2. **Cryptology area.** Digital signatures, Watermarking

1.3. **Keywords.** Medical imagery, scans, computed tomography, cryptography, computer vision, generative AI, medical image analysis

1.4. **Project description.** We aim to identify watermarking techniques that are the least obtrusive to computer vision and AI models meant for auto-analyzing Computed Tomography (CT) scans, Magnetic Resonance Imaging (MRI), and X-Ray images.

1.5. **Group member responsibilities.**

- Sean Moulton: Implementation of watermarking technique inspired by Guo and Zhuang [3], along with presentation.
- Anupreet Singh: Reproducing the results of Apostolidis and Papakostas [2] and providing writing assistannce.
- Omkar Kulkarni: Implementation of watermarking technique inspired by Rahimi and Rabbani [10], along with poster creation.

1.6. **Total budget.** $97,054

1.7. **Deliverables.**

1.7.1. *Progress Report.*

- Codebase showing Postolidis and Papakosta's [2] with findings reproduced.
- Written Report comparing reproduced results with the original findings.

1.7.2. *Final Report Draft.*

- Codebase with implementation of alternative watermarking methods.
- Written Report with progress updates, challenges encountered and deviations from initial approach (if any).
- Presentation and Poster highlighting the motivation and objectives achieved.

1.7.3. *Final Report.*

- Final Report showing analysis and summary of performance of deep learning computer vision models on all watermarking techniques implemented.
- Final codebase, presentation, and poster for showcasing the methodology, findings, and conclusions.

2. Executive Summary

2.1. **Project title.** Analyzing the effects of Alternative Watermarking Techniques on Deep Learning Computer Vision Tasks

2.2. **Date.** 11 February 2025

2.3. **Authors.**

(1) Sean Moulton, UMBC CMSC PhD student, seanmoulton@umbc.edu
(2) Anupreet Singh, UMBC CMSC Master's student, anuprel1@umbc.edu
(3) Omkar Kulkarni, UMBC CMSC PhD student, omkar.kulkarni@umbc.edu

2.4. **Project Keywords.** Medical Imagery, scans, public key cryptography, message authentication codes, generative AI

2.5. **Summary.** Recent developments in the field of artificial intelligence (especially generative AI) have made it easy to generate synthetic medical imagery such as CT scans [8]. While digital watermarking presents a promising solution for ensuring image authenticity, its effects on computer vision algorithms used in medical image analysis raise concern [2]. Many medical images are routinely processed by AI and computer vision models for critical tasks like diagnosis, organ segmentation, and anomaly detection. We propose to study how different watermarking algorithms affect the performance of these vision systems. Specifically, we want to analyze two modern watermarking approaches: a contourlet transform-based algorithm [10] and a difference expansion transform-based algorithm [3]. We will compare how these watermarking algorithms affect the accuracy of downstream computer vision tasks, aiming to identify which approach disrupts the accuracy of these tasks the least while maintaining security. To accomplish this, we will implement both watermarking algorithms on a dataset of CT scans and examine the outputs of various computer vision tasks such as contouring, object detection, and segmentation. We will use standardized metrics to measure both image similarity and vision task performance, creating a comprehensive framework for evaluating watermarking algorithms in the context of medical image analysis. Furthermore, we will suggest approaches to optimize watermarking for computer vision applications. We intend to write a detailed report comparing both watermarking algorithms with each other as well as our baseline. The budget for this work will involve funding three graduate students working as a team for one semester to cover their stipends at an expense of $97,054 when factoring indirect and direct factors.

3. Motivation

With the rise of artificial intelligence systems capable of generating synthetic medical imagery, the integrity of medical scans has become a critical concern. AI-generated CT scans have been shown to improve image detection models [7], but this same technology could be used to create falsified CT scans of real patients, leading to serious medical and ethical implications. To counter this, digital watermarking has emerged as a method to verify the authenticity of medical images.

Computer vision plays a crucial role in medical image analysis, aiding in the detection, classification, and diagnosis of diseases. Again, the work by Mangalagiri et al. [7] shows

a direct application of computer vision in medical image analysis on the detection of COVID-19 with a computer vision model. However, these models rely heavily on the integrity of their input data and can be fooled by trivial changes such as what is seen in the one pixel attack paper [11]. Understanding what kinds of modifications to the inputs can cause these kinds of undesirable effects and why they cause those effects is important to understand.

Apostolidis and Papkostas [2] have explored the effects of watermarking on medical image analysis. Their findings indicate that digital watermarking can be used to degrade the accuracy of deep learning based computer vision algorithms used in medical image analysis. However, their analysis is limited to a single algorithm. Understanding the effects that these techniques can have on deep learning based computer vision algorithms is crucial. By analyzing multiple watermarking algorithms and their impact on the algorithms used in medical image analysis, wee seek to provide a deeper understanding of the trade-offs between security, image quality, and algorithmic performance.

4. Previous Work

Various researchers have focused on exploring digital watermarking techniques in images and other media. Some of the earliest research we explored was from Zhou, Huang, and Lou [13], who presented a method to verify the authenticity and integrity of digital mammography. Their work, while having potential security issues today [12], laid the groundwork for future exploration. Jian Ren and Tongtong Li [5] proposed a computationally efficient cryptographic watermarking technique using signal processing techniques. Kuang, Zhang, and Han [6] proposed a system for authenticating medical images using reversible digital watermarking based on the RSA cryptosystem. This method proves to be effective, but public access to data proves to be a logistic issue.

With the advent of new AI-driven content generation technologies, it is now easier than ever for anyone to generate fake medical imagery. A study by Mangalagiri et al. [7] presents a method to improve computer vision algorithms using AI-generated images. This study shows promising results for influencing the accuracy of computer vision algorithms. The need for watermarking algorithms for verifying authenticity is clear. The work presented by Apostolidis and Papakostas in 2022 [2] shows another interesting angle. Their work shows that digital watermarking can be used as a way to degrade the accuracy of deep learning based computer vision algorithms used in medical image analysis. Understanding the effects that these techniques can have on deep learning based computer vision algorithms is crucial.

Numerous watermarking algorithms have been proposed and examined in the literature [9, 4, 1, 5, 6, 3, 10]. For our comparative analysis, we focus on implementing two specific approaches: Rahimi and Rabbani [10] introduce a contourlet transform-based algorithm, Guo and Zhuang [3] propose a difference expansion transform-based algorithm.

5. Specific Aims

Our project aims to analyze the effect that watermarking algorithms have on computer vision based algorithms used in medical image analysis. We aim to build on the existing

work by Apostolidis and Papakostas [2] by expanding their analysis to a broader array of watermarking algorithms. The algorithms we have chosen are an algorithm proposed by Rahimi and Rabbani [10] introduce a contourlet transform-based algorithm, and an algorithm by Guo and Zhuang [3] that proposes a difference expansion transform-based algorithm. These two algorithms, along with the algorithm originally used by Apostolidis and Papakostas [1] represent a range of different approaches used in digital watermarking. From the analysis that we do on these algorithms, we aim to generate conclusions about the nature that digital watermarking has on medical image analysis.

6. Plan

The aim of the project is to comprehensively compare watermarking algorithms and their effects on downstream computer vision and AI tasks. From this analysis, we aim to find the watermarking algorithm that has the least effect on the accuracy of the computer vision and AI algorithms used. With this in mind, we have organized the execution of our project into three primary phases.

6.1. **Phase 1 (Weeks 1 - 4).** We will reimplement the paper by Apostolidis and Papakostas [2] with our own dataset. This paper uses the research by Ali et al. [1] as the baseline watermarking algorithm, which we will implement on a computed tomography (CT) scan dataset of our own. We will then use similar computer vision techniques to what was outlined in the Apostolidis and Papakostas paper [2] and compare the results with our dataset to the findings of the authors. We anticipate that the general trend of the results should not vary with the changes to the dataset.

6.2. **Phase 2 (Weeks 4 - 8).** We implement the watermarking algorithms detailed by Farhad Rahimi and Hossein Rabbani, and Kyriakos D. Apostolidis and George A. Papakostas [10, 3], and apply it to watermark the CT scans in our dataset. We will use standardized metrics to measure how similar (or different) the watermarked and original images are. We will then run our suite of computer vision tasks on the newly watermarked images to generate a new set of metrics.

6.3. **Phase 3 (Weeks 8 - 12).** We comprehensively analyze the output metrics against the baseline expectations presented in phase 1. Using standard evaluation methods, we will compare the effect each cryptographic algorithm had on the performance of the computer vision tasks. Using this information, we will attempt to draw conclusions and form hypothesis about how cryptographic watermarking algorithms affect deep learning based computer vision algorithms.

7. Deliverables

7.1. **Progress Report.**

- Codebase showing Postolidis and Papakosta's [2] findings reproduced.
- Written Report comparing reproduced results with the original findings.

7.2. **Final Report Draft.**

- Codebase with implementation of alternative watermarking methods.
- Written Report with progress updates, challenges encountered and deviations from initial approach (if any).
- Presentation and Poster highlighting the motivation and objectives achieved.

7.3. **Final Report.**

- Final Report showing analysis and summary of performance of deep learning Computer Vision(CV) Models on all watermarking techniques implemented.
- Final Codebase, Presentation, and Poster for showcasing the methodology, findings, and conclusions.

8. Issues

Some of the foreseeable issues for this project could include:

- **Incompatibility of Watermarking Techniques with Data from [2]:** The two new watermarking techniques may not be fully compatible with the dataset used in Paper [2], which could affect the validity of the comparison and the ability to reproduce the findings accurately.
- **Analyzing and Comparing Volumetric data:** Volumetric data has three dimensions, which means the analysis and comparison needs to be done for all slices of the scan. This is a hard problem as metrics are calculated across three dimensions, and is also computationally expensive.

9. Broader Impact

Dependence on reliable AI medical analysis systems can have a huge impact on public health by alleviating the workload of overworked human professionals and enabling faster care for a larger number of patients. But, this will only be plausible if the AI analysis systems are robust in performing their analysis for authentic medical images like CT scans, MRI's and X-Rays images achieved through watermarking techniques. Hence, it is important to identify which watermarking technique has the best collaborative output with autonomous analysis systems while maintaining the provenance and authenticity is a key step in helping out society in the future. Our research aims to push forward in this direction by comparing three varied watermarking techniques on three varied medical image analysis deep learning systems to identify which combination results in the most efficient and robust system.

10. Budget

This will be a semester-long project, involving three graduate students working 20 hours/week. The graduate stipend level is 1111.56$ biweekly. One semester is nine such pay periods. Thus, the total stipend pay will be $1111.56 * 9 * 3 = 30,012$\$ University tuition and health insurance will be 742$ per credit and 2,738$ per student per semester A basic laptop for each student will be up to 1,200$ which adds another 3,600$. We will run AI inference on our data, which will require the use of GPUs. We estimate 1 week of GPU inference time, which is estimated to cost 750$. The exact costs are outlined in table 1 below.

Expense	Amount and/or Quantity	Totals
Graduate student stipend	1111.56$ biweekly	-
Biweekly pay periods	9	-
Number of students	3	-
Total graduate stipend	10,004$/3	30,012$
Graduate student tuition	742$/credit	-
Max credits	9	-
Total graduate tuition	6,678$/3	20,034$
Health insurance	2,738$/3	8,214$
Laptop	1,200$/3	3,600$
GPU server	4.5$ per hour/168 hours	756$
Total cost	-	62,616$

Table 1. Estimated budget

The university overhead is 55%, thus our total budget is 97,054$.

11. Bibliography

[1] Musrrat Ali et al. "An Optimized Digital Watermarking Scheme Based on Invariant DC Coefficients in Spatial Domain". In: *Electronics* 9.9 (Sept. 2020). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 1428. ISSN: 2079-9292. DOI: 10.3390/electronics9091428. URL: https://www.mdpi.com/2079-9292/9/9/1428 (visited on 02/10/2025).

[2] Kyriakos D. Apostolidis and George A. Papakostas. "Digital Watermarking as an Adversarial Attack on Medical Image Analysis with Deep Learning". In: *Journal of Imaging* 8.6 (May 30, 2022), p. 155. ISSN: 2313-433X. DOI: 10.3390/jimaging8060155.

[3] Xiaotao Guo and Tian-ge Zhuang. "A Region-Based Lossless Watermarking Scheme for Enhancing Security of Medical Data". In: *Journal of Digital Imaging: the official journal of the Society for Computer Applications in Radiology* 22.1 (Feb. 2009), pp. 53–64. ISSN: 0897-1889. DOI: 10.1007/s10278-007-9043-6. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3043669/ (visited on 02/11/2025).

[4] Zunera Jalil and Anwar M. Mirza. "A Review of Digital Watermarking Techniques for Text Documents". In: *2009 International Conference on Information and Multimedia Technology*. 2009 International Conference on Information and Multimedia Technology. Dec. 2009, pp. 230–234. DOI: 10.1109/ICIMT.2009.11. URL: https://ieeexplore.ieee.org/document/5381212/?arnumber=5381212 (visited on 02/11/2025).

[5] Jian Ren and Tongtong Li. "A Cryptographically Secure Image Watermarking Scheme". In: *MILCOM 2005 - 2005 IEEE Military Communications Conference*. MILCOM 2005 - 2005 IEEE Military Communications Conference. Atlantic City, NJ, USA: IEEE, 2005, pp. 1–6. ISBN: 978-0-7803-9393-6. DOI: 10.1109/MILCOM.2005.1605812. URL: http://ieeexplore.ieee.org/document/1605812/ (visited on 02/10/2025).

[6] Li-Qun Kuang, Yuan Zhang, and Xie Han. "A Medical Image Authentication System Based on Reversible Digital Watermarking". In: *2009 First International Conference on Information Science and Engineering*. 2009 First International Conference on Information Science and Engineering. ISSN: 2160-1291. Dec. 2009, pp. 1047–1050. DOI: 10.1109/ICISE.2009.60. URL: https://ieeexplore.ieee.org/document/5455333 (visited on 02/10/2025).

[7] Jayalakshmi Mangalagiri et al. "Toward Generating Synthetic CT Volumes using a 3D-Conditional Generative Adversarial Network". In: *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*. 2020 International Conference on Computational Science and Computational Intelligence (CSCI). Dec. 2020, pp. 858–862. DOI: 10.1109/CSCI51800.2020.00160. URL: https://ieeexplore.ieee.org/abstract/document/9458065 (visited on 02/10/2025).

[8] Sumeet Menon et al. "CCS-GAN: COVID-19 CT Scan Generation and Classification with Very Few Positive Training Images". In: *Journal of Digital Imaging* 36.4 (Apr. 17, 2023), pp. 1376–1389. ISSN: 1618-727X. DOI: 10.1007/s10278-023-00811-

2. URL: https://link.springer.com/10.1007/s10278-023-00811-2 (visited on 02/10/2025).

[9] Seyed Mojtaba Mousavi, Alireza Naghsh, and S. A. R. Abu-Bakar. "Watermarking Techniques used in Medical Images: a Survey". In: *Journal of Digital Imaging* 27.6 (Dec. 2014), pp. 714–729. ISSN: 0897-1889. DOI: 10.1007/s10278-014-9700-5. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4391065/ (visited on 02/11/2025).

[10] Farhad Rahimi and Hossein Rabbani. "A dual adaptive watermarking scheme in contourlet domain for DICOM images". In: *BioMedical Engineering OnLine* 10.1 (June 17, 2011), p. 53. ISSN: 1475-925X. DOI: 10.1186/1475-925X-10-53. URL: https://doi.org/10.1186/1475-925X-10-53 (visited on 02/12/2025).

[11] Jiawei Su, Danilo Vasconcellos Vargas, and Sakurai Kouichi. "One pixel attack for fooling deep neural networks". In: *IEEE Transactions on Evolutionary Computation* 23.5 (Oct. 2019), pp. 828–841. ISSN: 1089-778X, 1089-778X, 1941-0026. DOI: 10.1109/TEVC.2019.2890858. arXiv: 1710.08864[cs]. URL: http://arxiv.org/abs/1710.08864 (visited on 02/12/2025).

[12] Xiaoyun Wang and Hongbo Yu. "How to Break MD5 and Other Hash Functions". In: *Advances in Cryptology – EUROCRYPT 2005*. Ed. by Ronald Cramer. Red. by David Hutchison et al. Vol. 3494. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 19–35. ISBN: 978-3-540-25910-7 978-3-540-32055-5. DOI: 10.1007/11426639_2. URL: http://link.springer.com/10.1007/11426639_2 (visited on 02/10/2025).

[13] X.Q. Zhou, H.K. Huang, and S.L. Lou. "Authenticity and integrity of digital mammography images". In: *IEEE Transactions on Medical Imaging* 20.8 (Aug. 2001). Conference Name: IEEE Transactions on Medical Imaging, pp. 784–791. ISSN: 1558-254X. DOI: 10.1109/42.938246. URL: https://ieeexplore.ieee.org/document/938246 (visited on 02/10/2025).