Applied Computing Journal https://doi.org/10.52098/acj.20255124
ISSN: 2788-9688 Volume 5, Issue 1, pp 372-385, January 2025
Received: 27/12/2024 Revised: 20/1/2025 Accepted: 10/3/2025

A Review of JPEG File Carving: Challenges, Techniques, and Future Directions

Maryam Al Zaabi[1], Aaisha S. AlShibli[1]
[1]University of Technology and Applied Science, Sohar, Oman
* Corresponding author: Maryam Al Zaab[1], Maryam.alzaabi@utas.edu.om

Abstract

JPEG file carving is an essential component of digital forensics, enabling the recovery of image files from storage devices where metadata is missing or corrupted. This review explores the evolution of JPEG file carving techniques, from traditional header/footer methods to advanced approaches leveraging machine learning, genetic algorithms, and hybrid systems. The study highlights the challenges associated with file fragmentation, metadata loss, and the complexities of modern storage systems, emphasizing the limitations of existing tools in addressing these issues. Emerging methodologies, such as entropy clustering, context-aware carving, and deep learning for automated validation, demonstrate significant potential for improving recovery accuracy and scalability. By examining these advancements, the review identifies critical research gaps and proposes future directions, including the development of real-time AI-based tools and standardized evaluation frameworks. The findings underscore the importance of continued innovation in JPEG file carving, ensuring that digital forensics remains effective in addressing the growing complexities of data recovery and cyber investigations.

***Keywords**:* digital forensics; entropy clustering; deep learning; image files; JPEG file carving.

1. Introduction

The JPEG file format, known for its balance of quality and compression, dominates digital imaging, playing a crucial role in personal, professional, and forensic contexts. From storing cherished memories to serving as evidence in cybercrime investigations, JPEG files are prevalent. Yet, recovering these files from fragmented or corrupted storage devices continues to challenge digital forensics experts worldwide (Ali & Mohamad, 2021).

In digital forensics, file carving is a vital technique used to retrieve data from storage devices when metadata is unavailable or inaccessible. This method is particularly significant in scenarios involving deleted or fragmented files, where the file system no longer provides reliable pointers to store data. JPEG file carving, however, introduces unique challenges due to the format's dependency on specific markers, internal structures, and compression methods. Successfully reconstructing these files can determine the outcome of critical investigations (Ali et al., a.2018).

Despite advances in carving techniques, the recovery of JPEG files from fragmented or corrupted storage is often incomplete or inaccurate. Fragmentation, where file segments are scattered across the disk, is especially problematic, as traditional tools struggle to identify and reassemble the pieces. Moreover, scenarios involving metadata loss, overlapping file fragments, or compressed file structures exacerbate these challenges. Current methods, while effective in simple cases, frequently fail in more complex scenarios, leaving a significant gap in forensic capabilities (Ali et al., b.2018). This review seeks to provide a comprehensive overview of existing JPEG file carving techniques, emphasizing their mechanisms, strengths, and limitations. Special attention is given to emerging methodologies, such as machine learning and genetic algorithms, which promise to address longstanding challenges in this field. By analyzing these approaches, this paper aims to identify research gaps and propose directions for future innovation in JPEG forensics (Ali & Mohamad, 2021).

2. Literature Survey

Online JPEG file carving has been extensively studied in digital forensics due to its critical role in recovering image data from fragmented or corrupted storage devices. Early approaches relied on traditional header/footer-based techniques, which identified JPEG files by detecting Start of Image (SOI) and End of Image (EOI) markers (González et al., 2024). Tools like Foremost and Scalpel (Ali & Mohamad, 2021) were effective in scenarios involving intact files, offering computationally efficient solutions for straightforward recovery tasks. However, their reliance on contiguous file storage meant these methods were inadequate for handling fragmented files, which are increasingly common in modern storage systems.

To address these limitations, structure-based carving techniques emerged, leveraging the internal organization of JPEG files. Methods utilizing Quantization Tables (DQT) and Huffman Tables (DHT) allowed tools like PhotoRec to validate and reconstruct partial files, even in cases of minor corruption (Mohamad et al., 2009). These approaches represented a significant improvement over traditional methods but still struggled with more severe fragmentation and computational inefficiencies, highlighting the need for further innovation.

The introduction of content-based techniques marked another significant advancement in JPEG file carving. By analyzing statistical properties such as entropy, these methods could identify and group JPEG fragments with greater precision. Studies demonstrated the effectiveness of entropy clustering in handling highly fragmented files, particularly in scenarios where metadata was unavailable (Sari & Mohamad, 2020). However, the computational demands of these techniques presented a new challenge, making them less feasible for large-scale or time-sensitive applications.

Recent advancements have focused on hybrid and AI-driven methods, which integrate multiple carving approaches to overcome traditional limitations. Hybrid techniques, such as those implemented in the RX_myKarve framework, combine structure-based and content-based methods with genetic algorithms to achieve higher recovery accuracy.

This framework demonstrated a 97% recovery rate on standardized forensic datasets, showcasing the potential of combining methodologies to address complex fragmentation scenarios (Ali et al., 2021). Additionally, the integration of machine learning, particularly convolutional neural networks (CNNs), has transformed JPEG file carving by automating fragment classification and reassembly. AI-driven methods not only enhance accuracy but also offer scalable solutions for handling large datasets (Sunitha et al., 2023).

Despite these advancements, significant gaps remain in the field. Scalability continues to be a critical issue, with many advanced methods struggling to efficiently process large volumes of data. The lack of standardized evaluation benchmarks further impedes the ability to compare and validate carving tools. Furthermore, the absence of real-time capabilities in most techniques limits their practical application in forensic investigations. Addressing these challenges will require continued innovation and collaboration across the digital forensics community to ensure that JPEG file carving remains an effective and reliable tool for data recovery.

3. Understanding the JPEG File Structure

The JPEG file format is a widely used standard for digital images, offering efficient compression with minimal quality loss. This efficiency is achieved through a specific structure and encoding process, which also makes JPEG files uniquely identifiable and challenging to reconstruct when fragmented or corrupted. This section is completely referenced from (Kou W., 2013).

3.1. Core Components of JPEG Files

JPEG files are structured with specific markers and segments, each serving a distinct function in encoding and decoding the image:

3.1.1 Markers

- JPEG markers are 2-byte codes that define the start and end of specific segments. For instance, the Start of Image (SOI) marker (FFD8) and the End of Image (EOI) marker (FFD9) denote the boundaries of the file.
- Other critical markers include Start of Frame (SOF) for defining compression parameters and Define Huffman Table (DHT) for encoding data.

3.1.2. Segments

Each marker introduces a segment containing specific metadata or image data. Examples include:

- Quantization Tables (DQT): Define how image data is compressed.
- Huffman Tables (DHT): Specify the entropy coding used in compression.
- Compressed Image Data: Encodes the pixel values after transformation and quantization.

3.2. Role of JPEG Compression in File Carving

JPEG compression involves converting the image into a series of coefficients using discrete cosine transform (DCT), quantizing these coefficients, and encoding them through Huffman coding. This process reduces file size but introduces dependencies between segments, complicating recovery when parts of the file are missing (Raid et al., 2014).

Forensic file carving tools rely heavily on these structural characteristics to locate and reconstruct JPEG files. However, these tools often encounter challenges when different scenarios are encountered such as (Pal & Memon, 2009):

- Headers or footers are missing, leaving files incomplete.
- Fragmentation disrupts the sequence of compressed data blocks.
- Compressed data overlaps with other files, leading to misinterpretation.

3.3. JPEG File Structure

Understanding the complexity of JPEG file structures is essential for developing robust carving techniques. Each marker and segment provide insights that can be used to identify, validate, and reassemble fragmented files. For instance:

- Identifying valid SOI and EOI markers help define file boundaries (Fernandez V, 2018).
- Analyzing DQT and DHT tables can confirm file integrity and enable reconstruction of corrupted sections (Mohamad et al., 2009).
- Leveraging unique statistical properties of JPEG data can aid in distinguishing fragments from non-JPEG data (Pal & Memon, 2009).

A comprehensive understanding of the JPEG file structure is essential for effective file carving. Each component, from markers to segments, provides critical clues for identifying, validating, and reconstructing files. This knowledge enhances the accuracy and reliability of recovery efforts, particularly in scenarios involving fragmentation or corruption, making it a cornerstone of successful forensic investigations.

3.4. Current JPEG File Carving Techniques

JPEG file carving is a critical aspect of digital forensics, aiming to recover JPEG images from storage media where file systems may be damaged, corrupted, or missing entirely. This process is vital in various forensic investigations, including cybercrime, legal evidence recovery, and data restoration (Sethi P, 2024). Unlike traditional file recovery, JPEG carving often relies on the structural and statistical properties of the JPEG format to identify, extract, and reconstruct files.

The evolution of JPEG carving techniques reflects advancements in both digital storage technology and forensic challenges. From simple header/footer identification to complex entropy-based approaches, these methods are designed to address issues like file fragmentation, metadata loss, and overlapping data. Each technique offers unique strengths and faces specific limitations (Hadi A, 2016). The distinguished techniques employed in JPEG file carving, detailing their processes, advantages, and challenges (Sari & Mohamad, 2020) are as follows:

3.4.1. Header/Footer-Based Carving

Header/footer-based carving remains the most basic yet widely used technique due to its simplicity and efficiency.

3.4.1.1. Process in Detail

- The tool scans the raw data of the storage medium for the SOI (FFD8) and EOI (FFD9) markers.
- Once located, the data between these markers is extracted and saved as a separate file.

3.4.1.2. Practical Applications

- Used for quickly recovering files from lightly fragmented storage media, such as USB drives or SSDs with minimal wear.
- For example, tools like Scalpel and Foremost utilize header/footer carving for JPEG recovery.

3.4.1.3. Challenges in Complex Scenarios

- Fragmented Data: When the SOI and EOI markers are separated by unrelated or corrupted data, these tools cannot link the segments effectively.
- Missing Markers: In cases where file headers or footers are overwritten, this method becomes unusable.

3.4.2. Structure-Based Carving

Structure-based carving dives deeper into the JPEG file's internal organization, exploiting the relationships between markers and segments.

3.4.2.1. Process in Detail

- The tool examines the internal structure of the file, including quantization tables (DQT) and Huffman tables (DHT), to validate the integrity of the file.
- Even if the file is fragmented, the presence of internal metadata can guide the reconstruction process.

3.4.2.2. Advantages Over Header/Footer Methods

- This method can identify and validate partial files. For example, if only the DQT or DHT segments are intact, the file can still be partially reconstructed.
- Tools like PhotoRec use structure-based methods to handle moderately corrupted files.

3.4.2.3. Limitations

- Requires intact metadata to function effectively.
- If fragmentation disrupts metadata, even this approach struggles.

3.4.3. Content-Based Carving

Content-based carving is one of the most sophisticated methods, leveraging the inherent statistical properties of JPEG compression to identify and reassemble file fragments.

3.4.3.1. Process in Detail

- JPEG files exhibit unique entropy patterns due to their compression algorithms.
- Data blocks are analyzed for statistical properties, such as high entropy levels indicative of compressed image data.
- Identified fragments are then grouped and reconstructed into complete files.

3.4.3.2. Strengths

- Excels in scenarios involving heavy fragmentation, where traditional methods fail.
- Does not depend on markers or metadata, making it suitable for highly corrupted files.

3.4.3.3. Real Word Example

Advanced tools like Adroit Photo Forensics incorporate entropy-based carving for JPEG recovery. These tools can recover files from severely damaged storage, such as hard drives affected by physical damage. Several challenges can be noted including:

- High computational demands limit their use in real-time forensic investigations.
- There is a risk of misidentifying data fragments, leading to false positives or incomplete files.

4. Hybrid Techniques

Hybrid approaches combine the strengths of multiple methods to achieve higher accuracy and reliability (Hadi A, 2016). Hybrid file carving techniques combine several methods to enhance file recovery precision and effectiveness, particularly in forensic examination. Hybrid techniques apply header/footer markers for initial identification, structure-based verification to maintain metadata integrity, and content-based carving to reassemble damaged files. A hybrid tool, for example, can first detect Start of Image (SOI) markers and then employ entropy analysis to reconstruct fragments of a lost file for full data recovery.

One of the main advantages of hybrid approaches is that they are able to deal with difficult recovery scenarios that both suffer from file fragmentation and corruption. They achieve high accuracy in recovery while maintaining computational efficiency and are, therefore, superior to previous carving techniques. In practical applications, hybrid carving methods are employed more and more in forensic software tools designed for next-generation storage devices like Solid-State Drives (SSDs) and cloud storage mediums. Their tendency to recover files depending on several aspects instead of purely relying on legacy file signatures has turned them into formidable tools of today's digital investigations. One of the most notable examples of a hybrid solution is Reassembly-Free Carving (RFC), which combines heuristic analysis and structural verification to rebuild files independent of conventional marker sequences. The method enhances data integrity and recovery time, and it is, therefore, a perfect solution for forensic analysis. Table1 provides an overall overview of the carving techniques.

Table1: an overview of carving techniques.

Techniques	**Strengths**	**Weaknesses**	**Applications**
Header/Footer -Based	Simple, fast, and effective for non-fragmented files	Fails with fragmented or corrupted files	USB drives, SSDs with minimal wear
Structure-Based	Leverages internal metadata for partial recovery	Requires intact metadata	Moderate corruption scenarios
Content-Based	Handles heavy fragmentation and metadata loss	Computationally intensive and risk of false positives	Severely damaged drives
Hybrid Techniques	Combines strengths of all approaches	Complexity in implementation	Complex modern forensic cases

As the complexity of storage devices and the frequency of file fragmentation increases, traditional methods like header/footer carving are becoming less effective (Ali et al., 2018). Advanced techniques such as content-based and hybrid methods provide better solutions for recovering fragmented or corrupted JPEG files (Hadi A, 2016). However, these methods demand higher computational resources, which underscores the need for further research and optimization in this field (Sari & Mohamad, 2020).

5. Challenges in JPEG Forensics

Psychology has a long history of devising creative strategies to measure the “unmeasurable,” whether the targeted variable is a mental process, an attitude, or the quality of teaching (e.g., Webb et al., 1966). In addition, psychologists have documented various heuristics and biases that contribute to the misinterpretation of quantitative data (Gilovich et al., 2002), including SET scores (Boysen, 2015a, 2015b; Boysen et al., 2014). These skills enable psychologists to offer multiple solutions to the challenge posed by the need to objectively evaluate the quality of teaching and the impact of teaching on student learning.

Recovery and examination of JPEG files in digital forensics is a complex task due to the characteristics of the JPEG file format and the evolving nature of storage technology. JPEG file carving is the recreation of files from damaged, fragmented, or corrupted environments where the usual recovery methods fail to function. While existing techniques provide solutions to most scenarios, they are beset with significant limitations (Sethi P, 2024). Modern

forensic analyses are faced with increasingly difficult challenges as storage systems grow more complex and digital crimes become more sophisticated. Issues such as file fragmentation, metadata corruption, and overlapping data fragments significantly impact recovery processes. Furthermore, the widespread use of encrypted and compressed storage systems and the computational intricacy of advanced carving methods pose more problems for forensic analysts (Kumar M, 2021). The following covers the significant technical, practical, and ethical challenges affecting JPEG forensics, highlighting gaps in current methodologies and their future research and tool development implications.

5.1. File Fragmentation

Fragmentation occurs when a file is stored in nonadjacent clusters on a storage device (Ali et al., a. 2018). Fragmentation also breaks the continuity of JPEG data, separating the Start of Image (SOI) marker from subsequent segments or the End of Image (EOI) marker (Ali et al., b. 2018). If interspersed with non-pertinent data, it is challenging to reassemble fragmented files (Hadi A, 2016). Consider a real-world Scenarios: Defragmented hard disks are usually highly fragmented, and recovery is more difficult. SSD deleted files, due to wear-level algorithms, are highly fragmented (Al-Hatmi & Yousif, 2017). Tool Limitations: Most forensic tools do not differentiate between fragments from different JPEG files (Sari & Mohamad, 2020).

5.2. Loss or Corruption of Metadata

JPEG files rely on metadata structures such as quantization and Huffman tables in order to decode them effectively (McKeown et al., 2018). Without the metadata, the carving tools cannot effectively decompress or validate the retrieved image (McKeown et al., 2018). Corrupted headers or missing markers typically render traditional recovery methods futile. Deliberate metadata removal is a common practice to circumvent forensic analysis in criminal investigations (Mohith S, 2023).

5.3. Overlapping Data Fragments

Overlapping fragments are prevalent in heavily utilized storage media (Lyle et al., 2022). Forensic software may mistakenly merge unrelated fragments, producing false positives or contaminated reconstructions. Overlapping data adds noise, rendering entropy-based or statistical carving methods difficult. For example, a single JPEG fragment may overlap with a text or video file, causing ambiguity during reconstruction.

5.4. Absence of Standardized Tools

Diversity in JPEG carving tools leads to inconsistency in recovery success rates (Ali et al., a. 2018). Tools vary in approach, with a tendency to specialize in some respects and overlook others. The absence of benchmark frameworks makes it challenging to compare recovery accuracy. Examiners may use multiple tools, making the process time-consuming and resource intensive.

5.5. Computational Resource Constraints

Advanced forensic techniques, such as entropy-based and hybrid techniques, require high computing resources. Big storage sizes take time to process with high-complexity algorithms (Mittal et al., 2021). Resource-intensive

techniques may not assist time-sensitive forensic analysis. The solution is as follows: GPU acceleration and parallel processing can increase efficiency (Mittal et al., 2021).

5.6. Encrypted and Compressed Storage

Modern storage devices are increasingly using encryption and compression for security and efficiency (Pal & Memon, 2009). Encrypted storage restricts raw data access without decryption keys. Compression algorithms alter data structure, and marker and segment identification are challenging. Emerging Concerns is Cloud storage services employ encryption and deduplication, which complicate JPEG carving further (Alshabibi et al., 2024).

5.7. Ethical and Legal Challenges

JPEG forensic analysis generally involves processing sensitive personal data, which is a legal and ethical issue (Nakade S, 2024). Data privacy needs to be protected during recovery, especially in non-criminal investigations. Legal constraints may hamper access to information in certain jurisdictions.

6. Emerging Approaches in JPEG File Carving

The challenges inherent in JPEG file carving have led to the development of innovative approaches aimed at enhancing accuracy, efficiency, and reliability. Emerging methods incorporate advanced algorithms, machine learning, and hybrid techniques to address the complexities of modern storage systems and fragmented data. These new techniques focus on overcoming traditional limitations while adapting to evolving forensic demands. Below, we examine some of the most prominent approaches, highlighting their methodologies, strengths, and applications in modern digital forensics (Sunitha et al., 2022).

6.1. Machine Learning and AI-Based Carving

Machine learning (ML) and artificial intelligence (AI) are increasingly applied in digital forensics, providing robust solutions for complex JPEG file recovery tasks (Ali et al., b. 2018). ML models are trained on large datasets of JPEG files to recognize patterns, identify fragments, and predict fragment boundaries. These models can handle overlapping fragments, metadata loss, and unconventional compression formats more effectively than traditional methods. Research projects have demonstrated the effectiveness of convolutional neural networks (CNNs) in identifying and reconstructing JPEG fragments in highly fragmented environments. AI-based tools can classify fragments based on their likelihood of belonging to a JPEG file. Also, AI algorithms can predict missing metadata, enabling partial reconstruction of files with damaged headers or footers.

6.2. Context-Aware Carving Techniques

Context-aware methods consider the logical structure and relationships between fragments within the storage medium (Sportiello & Zanero, 2012). These techniques analyze file systems, clusters, and block relationships to improve fragment matching. Context-aware approaches reduce false positives and improve recovery rates by incorporating knowledge of common storage patterns. These methods are useful in forensic investigations involving modern file systems like NTFS or EXT4, where data clusters are often interlinked.

6.3. Genetic Algorithms for Fragment Reassembly

Genetic algorithms (GAs), inspired by natural selection, have shown promise in solving the file reassembly challenge (Ali et al., 2023). GAs simulate evolution by creating populations of potential fragment arrangements and refining them iteratively based on a fitness function. The fitness function evaluates how well a given arrangement matches the expected JPEG structure. It is capable of handling highly fragmented and unordered data. In addition, it can adapt to various JPEG file types and compression levels. For example: Studies have used GAs to reassemble JPEG files from severely fragmented storage, achieving higher accuracy than traditional entropy-based methods.

6.4. Entropy Clustering and Statistical Modeling

Advanced statistical approaches identify JPEG fragments based on their entropy and compression characteristics (Sunitha et al., 2022). Fragments are grouped based on shared statistical properties, such as entropy levels indicative of JPEG compression. Models use probabilistic techniques to estimate the likelihood of fragments belonging to a specific file. Particularly effective for JPEG files that have undergone non-standard compression or partial encryption. Tools using these methods can handle scenarios with overlapping or noisy data.

6.5. Deep Learning for Automated Validation

Deep learning techniques have emerged as a reliable method for automating the validation of reconstructed JPEG files (Hussain et al., 2021). Trained neural networks assess the integrity of a reconstructed file by analyzing its structure, metadata, and visual patterns. These systems are less prone to human error and can operate on a scale. It used to reduce the time required for manual validation in large-scale forensic investigations. Also, it can detect anomalies or partial reconstructions that traditional tools might overlook.

6.6. Hybrid and Modular Approaches

Hybrid methods integrate multiple carving techniques to achieve better results in diverse forensic scenarios (Boiko et al., 2023). These approaches combine header/footer, structure-based, and content-based methods, along with machine learning or statistical models. Modular systems allow investigators to customize their approach based on specific challenges. It can increase flexibility and accuracy in handling fragmented or corrupted files. Also, it balances computational demands with recovery effectiveness.

In conclusion: Emerging approaches in JPEG file carving offer promising advancements by integrating machine learning, statistical modeling, and hybrid methods. These techniques address longstanding challenges, such as fragmentation and metadata loss, while paving the way for more automated and accurate forensic tools. Continued research and development in these areas will further enhance the capabilities of digital forensics, ensuring that investigators can keep pace with the growing complexity of data recovery.

7. Results and Discussion

The review of existing literature highlights the significant evolution of JPEG file carving techniques, reflecting advancements in digital forensics. Traditional methods, such as header/footer-based carving, have served as foundational approaches. These techniques excel in recovering non-fragmented files by identifying the Start of Image

(SOI) and End of Image (EOI) markers. However, their effectiveness diminishes significantly in fragmented or corrupted data scenarios, where file segments are scattered across the storage medium. The reliance on complete and intact markers makes these methods inadequate for handling modern forensic challenges.

To address these limitations, hybrid techniques have emerged as a promising solution by combining the strengths of multiple carving methods. For instance, the RX_myKarve framework integrates structure-based and content-based approaches with advanced machine learning algorithms, such as genetic algorithms, to improve accuracy in fragmented file recovery. Studies evaluating RX_myKarve have reported recovery rates as high as 97% in standardized datasets like DFRWS 2006 and 2007, demonstrating its effectiveness in handling complex fragmentation scenarios. Hybrid approaches not only enhance recovery rates but also provide a flexible framework for tackling diverse forensic challenges.

Further advancements in artificial intelligence (AI) have introduced machine learning-based techniques that significantly improve file carving efficiency and accuracy. Convolutional neural networks (CNNs) and other deep learning models have shown exceptional promise in automating fragment classification and reassembly, particularly in cases involving severe fragmentation and metadata loss (Alkishri et al., 2023; Alkishri et al., 2024). These models analyze visual patterns and statistical properties of JPEG fragments, enabling more precise reconstruction of files. Despite their success, the application of AI in file carving presents challenges, including high computational demands and the need for extensive training datasets. This limits their scalability and practical utility in time-sensitive forensic investigations.

While these advancements mark significant progress, critical challenges remain. The absence of standardized datasets and evaluation frameworks creates difficulties in comparing and validating carving techniques. Furthermore, most advanced methods are computationally intensive, posing scalability issues when processing large volumes of data. Real-time application is another area requiring improvement, as current methods often struggle to balance computational efficiency with accuracy. Addressing these limitations is crucial for ensuring that JPEG file carving techniques remain effective and adaptable to the growing complexity of digital storage systems. Future research should focus on developing lightweight algorithms, establishing universal benchmarks, and exploring scalable solutions to bridge these gaps. Table2 summarizing the results of JPEG file carving techniques.

Hasoon (Hasoon et al. 2011) came up with an improvement method with a nonlinear filtering-based neural network, where they demonstrated an application of improvement that effectively minimizes the noise and maintains key details, which thus comes into play in detecting JPEG format with minimal compression artefacts and evident forensic image analysis.

The word cloud visually represents key concepts related to JPEG file carving techniques in digital forensics as shown in Figure 1. Larger words indicate more frequently discussed or emphasized techniques, while smaller words represent less dominant but still relevant terms. "Carving" and "techniques" appear the largest, indicating that file carving is the central focus. "Machine", "context-aware", and "hybrid" also appear prominently, suggesting modern approaches that incorporate AI-based methods and advanced heuristics. "Structure-based" and "content-based" carving methods indicate that both traditional marker-based and data-driven approaches are significant in digital forensics. "Statistical" and "statistical-based" approaches suggest increasing reliance on probabilistic models for fragment recovery.

"Entropy" and "clustering" indicate the use of data analysis and pattern recognition to classify and reassemble fragmented JPEG files. "Genetic" (likely referring to Genetic Algorithms) suggests the evolutionary optimization approach to improve file recovery. "Learning-based" (likely referring to machine learning and deep learning) suggests the integration of AI in modern forensic techniques. "Hybrid" implies that combining multiple methods is becoming the gold standard in forensic carving.

Table 2: summarizing the results of JPEG file carving techniques.

Technique	Description	Key Benefits	Challenges
Header/Footer-Based Carving	Identifies files using Start of Image (SOI) and End of Image (EOI) markers.	Simple and efficient for non-fragmented files.	Ineffective for fragmented or corrupted files.
Hybrid Techniques (e.g., RX_myKarve)	Combines structure-based, content-based, and machine learning approaches.	High accuracy in handling fragmented files; adaptable.	Computationally intensive; requires advanced algorithms.
Machine Learning-Based Carving (CNNs, Deep Learning)	Uses trained neural networks to recognize and reconstruct JPEG fragments.	Can handle severe fragmentation and metadata loss.	Requires large datasets and high computational power.
Context-Aware Carving	Analyzes file systems, clusters, and logical relationships between data fragments.	Reduces false positives and improves recovery rates.	May not work well with encrypted or compressed storage.
Genetic Algorithms for Fragment Reassembly	Use evolutionary algorithms to determine the best arrangement of file fragments.	Effective for highly fragmented and unordered data.	Computationally expensive requires iterative optimization.
Entropy Clustering & Statistical Modeling	Groups fragments based on statistical properties like entropy and compression.	Works well for non-standard compression and partial encryption.	May struggle with noise and overlapping fragments.
Deep Learning for Automated Validation	Uses neural networks to validate recovered files based on structure and metadata.	Reduces manual effort and improves detection of anomalies.	High resource requirements; potential overfitting issues.

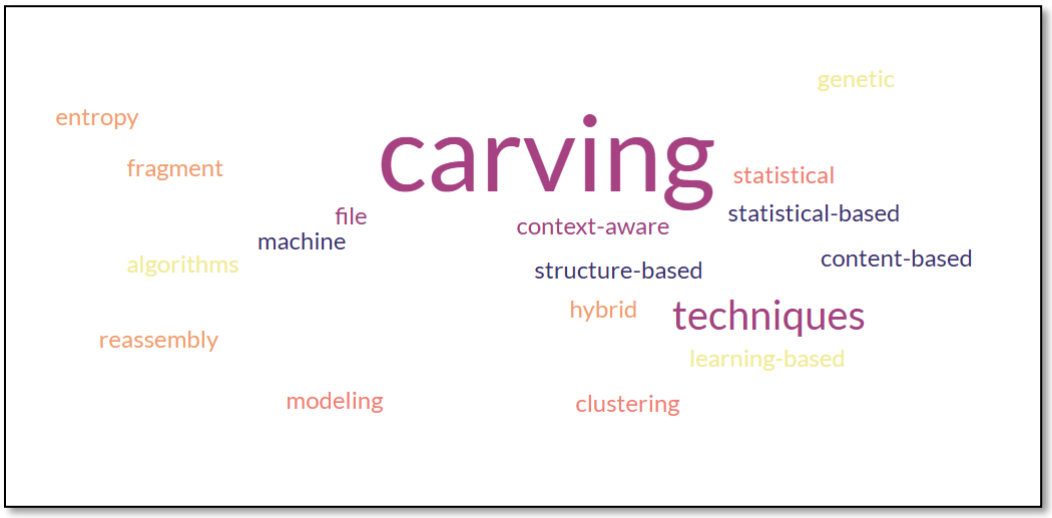

Figure 1: word cloud key concepts related to JPEG file carving techniques

8. Conclusion

JPEG file carving is a critical aspect of digital forensics, playing an indispensable role in recovering images from storage devices where traditional file recovery methods fail. As digital storage systems become increasingly complex and fragmented, the ability to retrieve JPEG files effectively has become essential in criminal investigations, cybersecurity, and legal evidence collection. Whether addressing accidental data loss or uncovering evidence in cybercrime, JPEG file carving is a foundational tool for forensic practitioners, bridging the gap between damaged data and actionable insights.

This review has examined a spectrum of techniques used in JPEG file carving, ranging from traditional header/footer methods to advanced machine learning and hybrid approaches. Traditional techniques, while straightforward and computationally efficient, often fall short in handling fragmented or corrupted files. Emerging methods, such as statistical modelling, entropy-based clustering, and genetic algorithms, offer promising advancements by addressing these limitations. Furthermore, hybrid techniques, which integrate multiple carving approaches, have shown significant potential in improving recovery accuracy and efficiency. Collectively, these advancements highlight the progress made in overcoming longstanding challenges like file fragmentation, metadata loss, and computational inefficiencies, paving the way for more reliable and automated solutions.

While significant progress has been made in JPEG file carving techniques, several areas remain ripe for further exploration and development. Future research should prioritize enhancing scalability to accommodate digital storage systems' increasing volume and complexity. The integration of artificial intelligence for real-time analysis is another promising avenue, enabling faster and more accurate recovery in time-sensitive forensic investigations. Additionally, establishing standardized datasets and evaluation frameworks is crucial for benchmarking and validating the performance of emerging techniques. By addressing these areas, the field of digital forensics can continue to evolve, ensuring investigators are equipped with the tools needed to meet the challenges of modern data recovery.

The advancements in JPEG file carving reflect the progress of digital forensics and underscore its critical role in addressing the complexities of modern data recovery. By combining innovation with practical application, these emerging techniques promise to enhance accuracy, efficiency, and reliability, ensuring forensic practitioners can keep pace with evolving storage technologies and digital crimes. As researchers and developers continue to push the boundaries of what is possible, the future of JPEG file carving holds the potential to revolutionize the field, strengthening its capacity to deliver justice in an increasingly digital world.

Acknowledgment

The research leading to these results has received no Research Grant Funding.

Author contribution: All authors have contributed, read, and agreed to the published version of the manuscript results.

Conflict of interest: The authors declare no conflict of interest.

References

[1]. Al-Hatmi, M. O., & Yousif, J. H. (2017). A review of Image Enhancement Systems and a case study of Salt &pepper noise removing. International Journal of Computation and Applied Sciences (IJOCAAS), 2(3), 171-176.

[2]. Alkishri, W., Widyarto, S., Yousif, J. H., & Al-Bahri, M. (2023). Fake face detection based on colour textual analysis using deep convolutional neural network. Journal of Internet Services and Information Security, 13(3), 143-155.

[3]. Alkishri, W., Widyarto, S., & Yousif, J. H. (2024). Evaluating the Effectiveness of a Gan Fingerprint Removal Approach in Fooling Deepfake Face Detection. Journal of Internet Services and Information Security (JISIS), 14(1), 85-103.

[4]. Ali, R. R., & Mohamad, K. M. (2021). RX_myKarve carving framework for reassembling complex fragmentations of JPEG images. Journal of King Saud University-Computer and Information Sciences, 33(1), 21-32.

[5]. Ali, R. R., Mohamad, K. M. B., Mostafa, S. A., Zebari, D. A., Jubair, M. A., & Alouane, M. T. H. (2023). A meta-heuristic method for reassemble bifragmented intertwined JPEG image files in digital forensic investigation. IEEE Access, 11, 111789-111800.

[6]. Ali, R. R., Mohamad, K. M., Jamel, S. A. P. I. E. E., & Khalid, S. K. A. a (2018). A review of digital forensics methods for JPEG file carving. J. Theor. Appl. Inf. Technol, 96(17), 5841-5856.

[7]. Ali, R. R., Mohamad, K. M., Jamel, S., & Khalid, S. K. A. b(2018). Extreme Learning Machine Classification of File Clusters for Evaluating Content-based Feature Vectors. International Journal of Engineering & Technology, 7(4.36), 167-171.

[8]. Alshabibi, M. M., Bu dookhi, A. K., & Hafizur Rahman, M. M. (2024). Forensic investigation, challenges, and issues of Cloud Data: A systematic literature review. Computers, 13(8), 213.

[9]. Boiko, M., Moskalenko, V., & Shovkoplias, O. (2023). Advanced file carving: ontology, models and methods.

[10]. Fernandez, V. (2018, November 13). Carving JPEGs from SOI to EOI. Retrieved from cyberphor: https://cyberphor.com/caringjpegs-from-soi-to-eoi

[11]. González Arias, R., Bermejo Higuera, J., Rainer Granados, J. J., Bermejo Higuera, J. R., & Sicilia Montalvo, J. A. (2024). Systematic Review: Anti-Forensic Computer Techniques. Applied Sciences, 14(12), 5302.

[12]. Hadi, A. (2016, August). Reviewing and evaluating existing file carving techniques for jpeg files. In 2016 Cybersecurity and Cyberforensics Conference (CCC) (pp. 55-59). IEEE.

[13]. Hasoon, F. N., Yousif, J. H., Hasson, N. N., & Ramli, A. R. (2011). Image enhancement using nonlinear filtering based neural network. Journal of Computing, 3(5), 171-176.

[14]. Hussain, I., Tana, S., Li, B., Xin, X., Hussain, D., & Huang, J. (2021). A Novel Deep Learning Framework for double JPEG Compression Detection of Small Size Blocks. Journal of Visual Communication and Image Representation, 1-10. 10.1016/j.jvcir.2021.103269

[15]. Kou, W. (2013). Digital image compression: algorithms and standards (Vol. 333). Springer Science & Business Media.

[16]. Kumar, M. (2021). Solid state drive forensics analysis—Challenges and recommendations. Concurrency and Computation: Practice and Experience, 33(24), e6442.

[17]. Lyle, J. R., Guttman, B., Butler, J., Sauerwein, K., Reed, C., & Lloyd, C. (2022). Digital investigation techniques: a NIST scientific foundation review.

[18]. McKeown, S., Russell, G., & Leimbach, P. (2018). Fingerprinting JPEGs with Optimised Huffman Tables. The Journal of Digital FORENSICS. INTERNATIONAL JOURNAL Forensics Security and Law, 1-15. OF CREATIVE RESEARCH THOUGHTS, 1-10.

[19]. Mittal, G., Korus, P., & Memon, N. (2021). FiFTy: Large-scale File Fragment TypeIdentification using Convolutional Neural Networks. Transactions on Information Forensics and Security, 28-41.

[20]. Mohamad, K. M., & Deris, M. M. (2009). Fragmentation Point Detection of JPEG Images at DHT Using Validator. Future Generation Information Technology, First International Conference (pp. 173-180). Jeju Island: ResearchGate.

[21]. Mohith, S. (2023, September 26). The Role of EXIF Data in Forensic Investigations. Retrieved from exifviewerapp.com: https://exifviewerapp.com/the-role-of-exifdata-in-forensic-investigations/

[22]. Nakade, S. I. (2024). PRIVACY CONCERNS AND ETHICAL ISSUES IN DIGITAL FORENSICS. International Research Journal of Modernization in Engineering Technology and Science, 428-436.

[23]. Pal, A., & Memon, N. (2009). The evolution of file carving. IEEE signal processing magazine, 26(2), 59-71.

[24]. Raid, A. M., Khedr, W. M., El-Dosuky, M. A., & Ahmed, W. (2014). Jpeg image compression using discrete cosine transform-A survey. arXiv preprint arXiv:1405.6147.

[25]. Sari, S. A., & Mohamad, K. M. (2020, May). A review of graph theoretic and weightage techniques in file carving. In Journal of Physics: Conference Series (Vol. 1529, No. 5, p. 052011). IOP Publishing.
[26]. Sethi, P. C. (2024). File Carving: Analyzing Data Retrieval in Digital, © 2024 IJCRT, Volume 12, Issue 4, pp555-564.
[27]. Sportiello, L., & Zanero, S. (2012). Context-based file block classification. In Advances in Digital Forensics VIII: 8th IFIP WG 11.9 International Conference on Digital Forensics, Pretoria, South Africa, January 3-5, 2012, Revised Selected Papers 8 (pp. 67-82). Springer Berlin Heidelberg.
[28]. Sunitha, M., Srinivas, K., Adilakshmi, T., Baswaraj, D., Perala, A., & Bellamkonda, V. (2022, December). Data Fragment Classification of High Entropy Files Using Machine Learning. In International Conference on Information and Management Engineering (pp. 627-633). Singapore: Springer Nature Singapore.
[29]. Sunitha, M., Srinivas, K., Adilakshmi, T., Baswaraj, D., Perala, A., & Bellamkonda, V. (2022, December). Data Fragment Classification of High Entropy Files Using Machine Learning. In International Conference on Information and Management Engineering (pp. 627-633). Singapore: Springer Nature Singapore.