

A method for detecting manipulated images in large-scale photo datasets based on histogram analytics and multi-color space analysis

Andriy Dudnik^{1,2,3,*}, Dmytro Zhuravlov^{2,*}, Andrii Fesenko^{4,*}, Vira Mykolaichuk^{1,*}, Vitalijus Cernej^{2,*}, Valentyn Datsko^{2,*} and Maksym Ivasiv^{2,*}

¹Kyiv National Taras Shevchenko University, Volodymyrska Str., 60, Kyiv, 03022, Ukraine

²Interregional Academy of Personnel Management, Frometivska Str., 2, Kyiv, 03039, Ukraine

³Open University of Human Development "Ukraine", Lvivs'ka Str., 23, Kyiv, 04071, Ukraine

⁴State Scientific and Research Institute of Cybersecurity Technologies and Information Protection, Maksym Zaliznyak Str., 3, Kyiv, 03142, Ukraine

Abstract

The rapid development of artificial intelligence, image editing software, and synthetic content generation technologies has significantly increased the amount of manipulated visual data in large-scale photo datasets. This creates a growing need for automated methods capable of detecting image modifications and ensuring digital content integrity. This paper proposes a method for detecting manipulated images in large-scale photo datasets based on histogram analytics and multi-color space analysis. The proposed approach combines histogram similarity assessment with the evaluation of image quality characteristics, including sharpness, contrast, noise level, color distribution, and local artifacts. To improve detection reliability, image analysis is simultaneously performed in RGB, YCrCb, and CIE Luv* color spaces, enabling the identification of inconsistencies introduced by image editing, photomontage, and AI-generated content. An integrated manipulation score is calculated using histogram deviations and extracted image features. The proposed method was implemented as a web-based software system and evaluated using authentic and manipulated images from publicly available datasets. Experimental results demonstrate that the method effectively detects image manipulations while maintaining low computational complexity and suitability for large-scale image analysis.

Keywords

image manipulation detection, histogram analytics, color space analysis, digital forensics, computer vision, image integrity verification, large-scale photo datasets, deepfake detection

1. Introduction

The rapid growth of digital technologies, cloud storage platforms, social media services, and artificial intelligence systems has led to an unprecedented increase in the volume of visual information generated and exchanged worldwide. Billions of photographs are uploaded, shared, and archived every day, forming large-scale photo datasets that are used in social networks, digital archives, news agencies, e-commerce platforms, surveillance systems, and scientific repositories. At the same time, the availability of advanced image editing software and AI-based content generation tools has significantly simplified the creation of manipulated visual content.

Modern image manipulation techniques include traditional photo editing, image compositing, object insertion and removal, style transfer, deepfake generation, and diffusion-based image synthesis. Recent advances in generative artificial intelligence enable the production of highly realistic images that are

WCH'26: Fifth International Workshop on Cyber Hygiene, June 25–26, 2026, Krakow, Poland

*Corresponding author.

†These authors contributed equally.

✉ a.s.dudnik@gmail.com (A. Dudnik); zhuravlyovdima34@gmail.com (D. Zhuravlov); andrii.fesenko@npp.kai.edu.ua (A. Fesenko); viramykolaichuk@knu.ua (V. Mykolaichuk); cernejvitalijus@gmail.com (V. Cernej); datsko1212@gmail.com (V. Datsko); maks1404@gmail.com (M. Ivasiv)

ORCID 0000-0003-1339-7820 (A. Dudnik); 0000-0003-0353-0544 (D. Zhuravlov); 0000-0001-5154-5324 (A. Fesenko); 0000-0002-2532-5771 (V. Mykolaichuk); 0000-0002-3725-6216 (V. Cernej); 0009-0006-7161-5039 (V. Datsko); 0009-0001-0205-0963 (M. Ivasiv)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

often difficult to distinguish from authentic photographs through visual inspection alone. As a result, manipulated images have become a serious challenge for digital forensics, information security, media verification, and automated content moderation systems.

The problem becomes particularly significant in large-scale photo datasets, where manual verification of image authenticity is impractical due to the enormous volume of data. Automated methods capable of identifying suspicious or manipulated images are therefore required to support efficient dataset validation and integrity assessment. Such methods are important not only for detecting intentionally forged content but also for identifying images affected by excessive post-processing, synthetic generation, or unauthorized modifications.

Existing approaches for image manipulation detection are primarily based on deep learning techniques, including convolutional neural networks, transformer architectures, and hybrid computer vision models. Although these methods often achieve high detection accuracy, they typically require extensive computational resources, large labeled datasets, and significant training time. Furthermore, their performance may degrade when applied to previously unseen manipulation techniques or heavily compressed images.

An alternative approach is based on the analysis of intrinsic image characteristics that are affected during the modification process. Image editing operations frequently introduce detectable changes in histogram distributions, color relationships, local texture properties, noise patterns, contrast levels, and edge structures. These modifications can be analyzed using classical computer vision techniques without the need for computationally expensive model training.

In addition to image authenticity verification, the ability to automatically detect manipulated images plays an important role in maintaining the reliability of machine learning datasets. Large image collections are frequently used for training and evaluating computer vision models, including object detection, image classification, biometric recognition, visual search systems, and content recommendation platforms. The presence of manipulated, duplicated, or synthetically generated images within such datasets may introduce biases, reduce model generalization capabilities, and negatively affect the performance of artificial intelligence applications. Consequently, efficient methods for identifying suspicious visual content are becoming an essential component of modern data quality assurance processes.

Another important aspect is the interpretability of manipulation detection results. Unlike many deep learning approaches that often operate as black-box systems, feature-based methods provide explicit information regarding the image characteristics that contributed to the final decision. The analysis of histogram distributions, color-space inconsistencies, noise patterns, and image quality indicators allows investigators and analysts to understand the potential reasons for image manipulation. Such transparency is particularly valuable in digital forensic investigations, legal proceedings, scientific data validation, journalism, and other application domains where explainability and reproducibility of results are critical requirements.

Furthermore, the growing use of generative artificial intelligence in image creation has introduced new challenges for content verification. Modern diffusion models and generative adversarial networks are capable of producing synthetic images with a high level of visual realism, often preserving natural textures, lighting conditions, and color distributions. As a result, conventional visual inspection methods are becoming increasingly ineffective. This situation highlights the necessity of developing computational approaches capable of identifying subtle statistical inconsistencies that remain present even in highly realistic synthetic or manipulated images.

This paper proposes a method for detecting manipulated images in large-scale photo datasets based on histogram analytics and multi-color space analysis. The proposed approach combines histogram similarity assessment with the evaluation of image quality characteristics, including sharpness, contrast, noise level, color distribution, and local artifacts. To improve detection reliability, image analysis is performed simultaneously in RGB , $YCrCb$, and $CIE L^*u^*v^*$ color spaces, enabling the identification of inconsistencies caused by image editing and synthetic content generation.

The scientific contribution of the study lies in the development of an integrated image manipulation detection method that combines multiple image quality metrics and color-space representations into a unified manipulation assessment framework suitable for large-scale image dataset analysis. The

practical significance of the proposed solution is its applicability to digital forensic investigations, image integrity verification systems, multimedia content monitoring platforms, and automated screening of large visual data collections.

2. Review of existing solutions and literature sources

The rapid development of image editing software, deep learning algorithms, and generative artificial intelligence has significantly increased the amount of manipulated visual content distributed through digital platforms. As a result, the detection of image manipulations has become an important research area in digital forensics, computer vision, and information security. Existing solutions can generally be divided into deep learning-based approaches, forensic feature analysis methods, camera fingerprinting techniques, and hybrid systems that combine multiple detection strategies.

One of the most influential contributions in this field is the FaceForensics++ framework proposed by Rössler et al. [1]. The authors introduced a large-scale benchmark dataset containing manipulated facial images generated using DeepFakes, FaceSwap, Face2Face, and NeuralTextures techniques. Their experiments demonstrated that convolutional neural networks can effectively detect image manipulations when trained on sufficiently large datasets. However, the approach requires substantial computational resources and extensive training procedures.

A lightweight alternative was proposed by Afchar et al. through the MesoNet architecture [2]. The method focuses on mesoscopic image features and achieves high detection performance while maintaining relatively low computational complexity. Nevertheless, the model is specifically optimized for facial manipulations and its applicability to heterogeneous photo collections remains limited.

Bayar and Stamm [3] introduced a constrained convolutional neural network designed to suppress image content and emphasize manipulation traces. Their work demonstrated that low-level image artifacts often contain more forensic information than semantic image content. This idea subsequently influenced many modern image forensic systems and highlighted the importance of intrinsic image characteristics for manipulation detection.

Another significant development was presented by Wu et al. [4], who proposed ManTra-Net, a deep neural network capable of detecting and localizing multiple types of image manipulations, including copy-move, image splicing, object removal, and enhancement operations. Although the method achieves high detection accuracy, its computational complexity makes large-scale image screening challenging.

The importance of acquisition-related image artifacts was investigated by Cozzolino and Verdoliva [5], who developed the Noiseprint framework. Their method extracts camera-model fingerprints using convolutional neural networks and allows the identification of inconsistencies introduced during image tampering. Similar studies demonstrated that camera acquisition traces can provide valuable forensic information even when visual manipulation artifacts are difficult to observe directly.

Comprehensive reviews presented in [6, 7, 8] indicate that modern manipulation detection systems increasingly rely on deep neural networks, attention mechanisms, and multimodal feature extraction. These studies also emphasize several unresolved problems, including sensitivity to image compression, reduced generalization to previously unseen manipulation techniques, vulnerability to adversarial attacks, and high computational requirements.

Recent research has focused on the detection of synthetic images generated by generative adversarial networks (GANs) and diffusion models. Wang et al. [9] demonstrated that CNN-generated images contain characteristic artifacts that can be exploited for forensic analysis. Similar observations were reported by Tariq et al. [10], who investigated the detection of machine-generated face images in uncontrolled environments. Marra et al. [11] further showed that GAN-generated images often leave artificial fingerprints that differ from those found in authentic photographs.

Subsequent studies expanded this concept by investigating image artifacts introduced by modern image generation pipelines. Gragnaniello et al. [12] demonstrated that GAN-generated images preserve detectable forensic traces even after distribution through social media platforms. Durall et al. [13] analyzed frequency-domain characteristics of synthetic images and showed that generative models

frequently fail to reproduce natural spectral distributions. Frequency-based detection approaches were further investigated by Frank et al. [14], who demonstrated the effectiveness of spectral analysis for identifying deepfake images.

The emergence of diffusion-based image generation systems has introduced new challenges for digital image forensics. Corvi et al. [15] analyzed the forensic properties of images generated by both GAN and diffusion models and reported that synthetic content still contains detectable statistical inconsistencies despite its increasing visual realism. Similar conclusions were reached by Dang et al. [16], who investigated the detection of digital face manipulations using forensic feature analysis. Li et al. [17] proposed the use of physiological inconsistencies, such as abnormal eye blinking patterns, for exposing manipulated facial content, while Guarnera et al. [18] demonstrated that convolutional traces can be successfully exploited for deepfake detection.

Despite the significant progress achieved by deep learning approaches, many existing solutions remain computationally expensive and require large annotated datasets for training. This limitation becomes particularly important when processing large-scale photo collections containing thousands or millions of images. In such scenarios, lightweight methods based on intrinsic image characteristics remain highly relevant. Histogram distributions, color consistency, noise patterns, contrast variations, sharpness measures, and local artifacts can often reveal manipulation traces without requiring complex neural network architectures. Based on the conducted literature review, it can be concluded that current research primarily focuses on deep learning-based image forgery detection, while comparatively less attention is paid to computationally efficient methods suitable for large-scale dataset analysis. Therefore, this study proposes a method based on histogram analytics and multi-color space analysis that combines histogram similarity assessment, image quality metrics, and color-space transformations into a unified framework for detecting manipulated images in large-scale photo datasets.

3. Problem formulation for manipulated image detection in large-scale photo datasets

The continuous growth of digital image repositories, cloud storage services, social media platforms, and artificial intelligence applications has resulted in the formation of large-scale photo datasets containing millions of visual records. These datasets are widely used in computer vision, machine learning, multimedia retrieval, digital archiving, surveillance systems, journalism, and social networking services. At the same time, the increasing accessibility of image editing tools and AI-based content generation technologies has significantly increased the number of manipulated images distributed within such collections.

Modern image manipulations include classical photo editing operations, image splicing, object insertion and removal, retouching, color correction, synthetic image generation using Generative Adversarial Networks (GANs), and diffusion-based image synthesis. Many of these modifications leave only subtle traces that are difficult to identify through visual inspection. Consequently, manual verification becomes ineffective when dealing with large-scale photo datasets.

Existing deep learning approaches provide high detection accuracy but often require extensive computational resources, large labeled training datasets, and long training procedures [1, 4, 6, 19]. Furthermore, their performance may decrease when confronted with previously unseen manipulation techniques, highly compressed images, or images originating from different acquisition devices [7, 15]. These limitations make it difficult to apply such approaches for rapid preliminary screening of large image repositories.

In contrast, image forensic methods based on the analysis of intrinsic image characteristics can be implemented with significantly lower computational complexity. Image manipulations frequently affect histogram distributions, color consistency, local texture properties, noise patterns, edge structures, and contrast relationships. Such changes can be quantified using classical image processing techniques and used as indicators of potential image tampering [3, 5, 11].

Let the analyzed photo dataset be represented as

$$D = \{I_1, I_2, \dots, I_n\}, \quad (1)$$

where I_i denotes an individual digital image and n represents the total number of images in the dataset.

For each image I_i , a feature vector is extracted:

$$F_i = \{H_i, R_i, C_i, N_i, A_i\}, \quad (2)$$

where H_i is a set of histogram descriptors extracted from image I_i , R_i denotes the sharpness feature, C_i characterizes image contrast, N_i corresponds to the estimated noise level, and A_i represents a local artifact measure used to identify potential manipulation traces. Together, these parameters form a multidimensional feature vector describing the visual and statistical properties of the analyzed image [20, 21, 22].

To improve robustness against different image processing operations, the feature extraction process is performed simultaneously in multiple color spaces:

$$CS = \{RGB, YCrCb, CIE L^*u^*v^*\}, \quad (3)$$

where RGB represents the standard additive color model used for digital image representation, $YCrCb$ denotes a luminance–chrominance color space separating brightness information from color components, and $CIE L^*u^*v^*$ is a perceptually uniform color space designed to approximate human color perception. The simultaneous use of these color spaces enables the extraction of complementary image characteristics and improves the detection of color inconsistencies, processing artifacts, and manipulation traces that may not be visible in a single color representation [23, 24].

The use of multiple color representations enables the detection of manipulation traces that may be hidden in one color space but become observable after transformation into another representation.

The main objective of this research is to develop a computationally efficient method capable of automatically identifying manipulated images within large-scale photo datasets. The method should combine histogram analytics, multi-color space analysis, and image quality assessment metrics into a unified framework suitable for large-scale image screening [25, 26, 27].

Accordingly, the research problem can be formulated as follows: given a large collection of digital images, it is necessary to determine whether an image contains signs of manipulation by analyzing its histogram characteristics, color-space consistency, sharpness, contrast, noise level, and local artifacts while minimizing computational complexity and preserving applicability to large datasets.

To solve this problem, an integrated manipulation score K_i is introduced. This score aggregates multiple image quality indicators and forensic features into a single numerical value describing the probability of image manipulation. Images with high values of K_i are considered suspicious and may be selected for further detailed forensic examination [28, 29].

Thus, the proposed approach aims to bridge the gap between computationally expensive deep learning solutions and lightweight forensic analysis techniques by providing an efficient method for manipulated image detection in large-scale photo datasets [30, 31].

4. Method for detecting manipulated images in large-scale photo sets

The proposed method is intended for the automated detection of manipulated images within large-scale photo collections. Unlike deep learning approaches that require extensive training datasets and significant computational resources, the proposed solution is based on the analysis of intrinsic image characteristics that are affected during image editing, synthetic image generation, compression, and other manipulation processes. The method combines histogram analytics, multi-color space analysis, and image quality assessment into a unified framework for image integrity verification.

The general workflow of the proposed method consists of five main stages: image acquisition and preprocessing, histogram analysis, multi-color space feature extraction, image quality assessment, and integrated manipulation score calculation (Figure 1).

According to formulas (1), (2), (3), each image is represented by a multidimensional feature vector extracted from several color space representations. These features form the basis of the manipulation detection procedure and are used to assess statistical and visual discrepancies within the analyzed image.

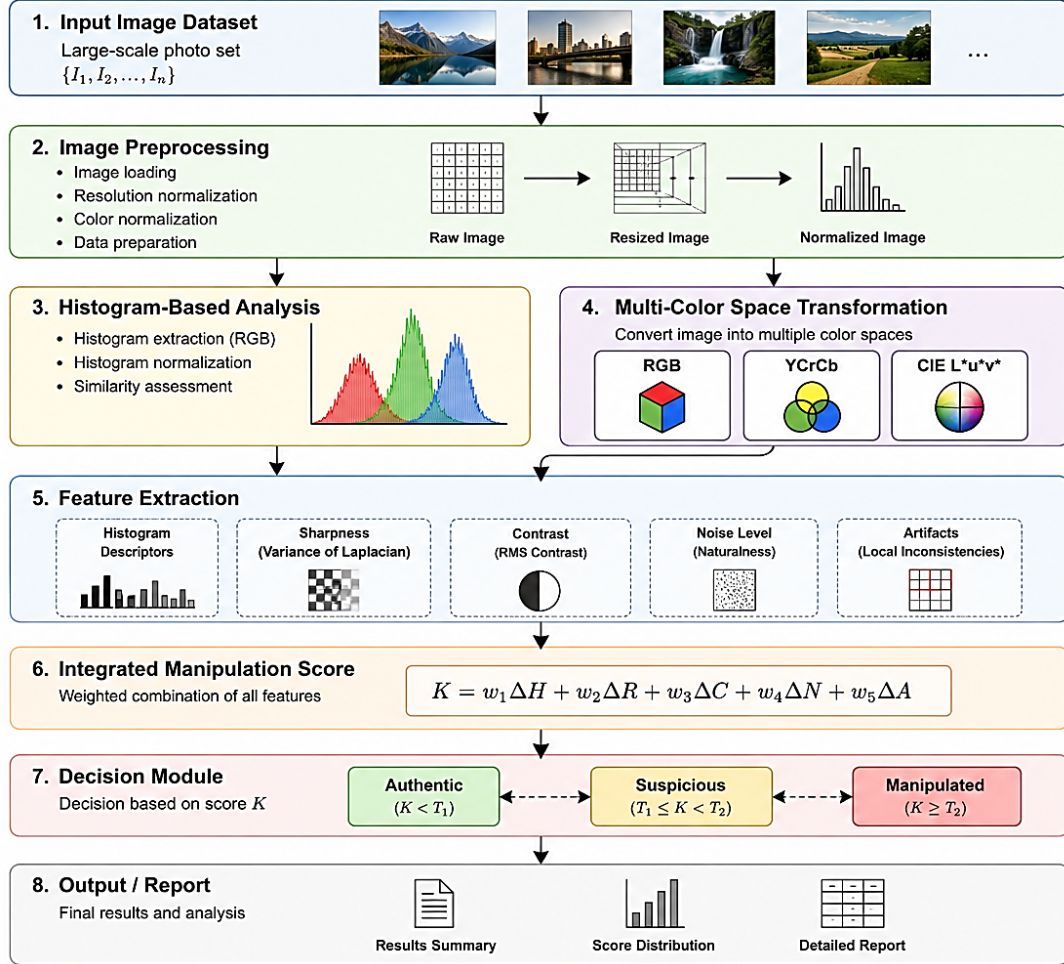


Figure 1: General workflow of the proposed image manipulation detection method.

The first stage of the proposed method is histogram-based analysis. Histogram distributions provide information about the statistical properties of pixel intensities and color values. Since image editing operations frequently alter brightness distributions, contrast relationships, and color balance, histogram deviations can be used as indicators of image manipulation. To quantify histogram similarity, the following measure is employed:

$$S = \frac{\sum_{i=1}^m H_1(i)H_2(i)}{\sqrt{\sum_{i=1}^m H_1(i)^2} \sqrt{\sum_{i=1}^m H_2(i)^2}} \quad (4)$$

where $H_1(i)$ and $H_2(i)$ denote the normalized histogram values of two analyzed images at the i -th histogram bin, while m represents the total number of histogram bins. The similarity score S ranges from 0 to 1, where values close to unity indicate highly similar histogram distributions (Figure 2).

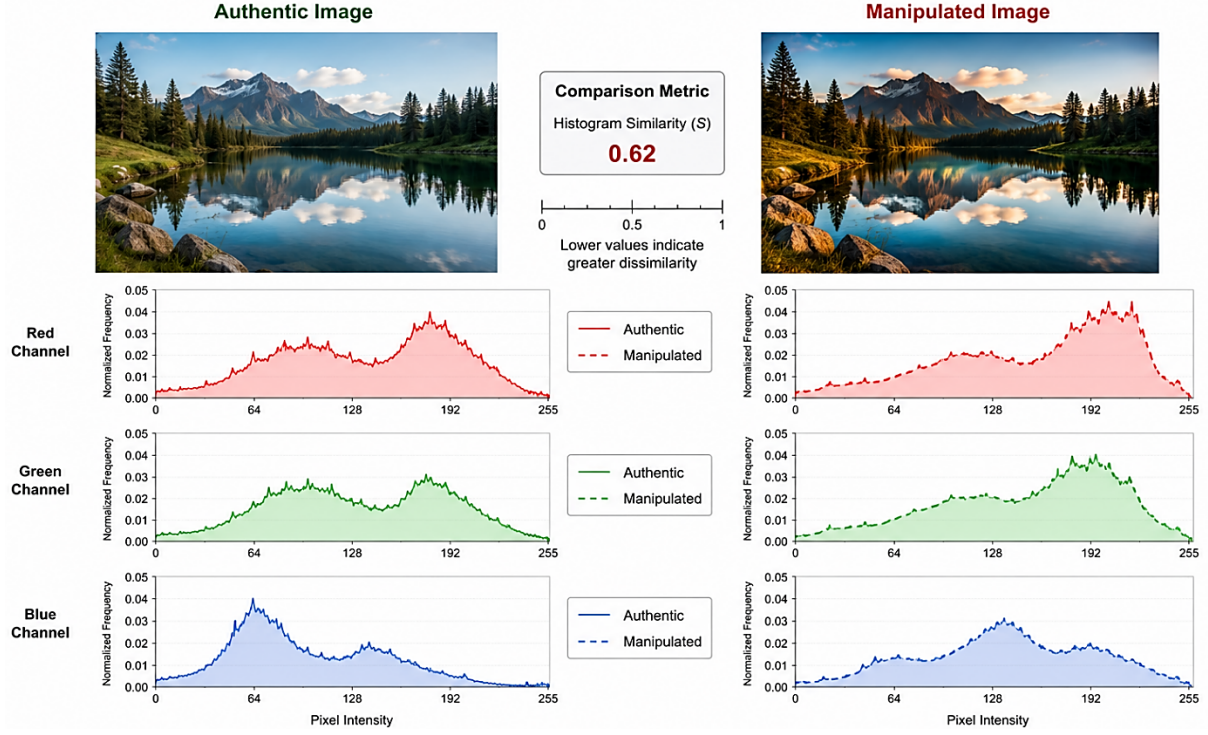


Figure 2: Example of histogram comparison between authentic and manipulated images.

To improve robustness against different manipulation techniques, the proposed method performs feature extraction in multiple color spaces. As described in Section 3, image analysis is simultaneously conducted in RGB , $YCrCb$, and $CIE L^*u^*v^*$ representations. Each color model emphasizes different image characteristics and therefore contributes complementary forensic information. The RGB model preserves the original color composition, $YCrCb$ separates luminance and chrominance information, while $CIE L^*u^*v^*$ provides a perceptually uniform representation of color differences. The combined use of these color spaces increases the probability of identifying manipulation traces that may remain hidden within a single color representation (Figure 3).

Following color-space transformation, several image quality metrics are calculated. These metrics characterize different visual properties of the image and provide additional information regarding possible manipulation traces. Image sharpness is estimated using the variance of the Laplacian operator:

$$R = Var(\nabla^2 I), \quad (5)$$

where $\nabla^2 I$ denotes the Laplacian-transformed image. Reduced sharpness values may indicate excessive smoothing, synthetic image generation, or aggressive image compression. Image contrast is evaluated using the root mean square contrast measure:

$$C = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (6)$$

where x_i represents pixel intensity values, $\bar{x}$ denotes the mean image intensity, and N corresponds to the total number of pixels. Significant deviations in contrast may indicate local modifications or artificial enhancement operations.

Noise level estimation is performed by analyzing local intensity variations within homogeneous image regions. Since image acquisition devices generate characteristic noise patterns, modifications introduced during editing or synthetic image generation often disrupt the natural noise structure. Consequently, abnormal noise distributions may indicate image tampering.

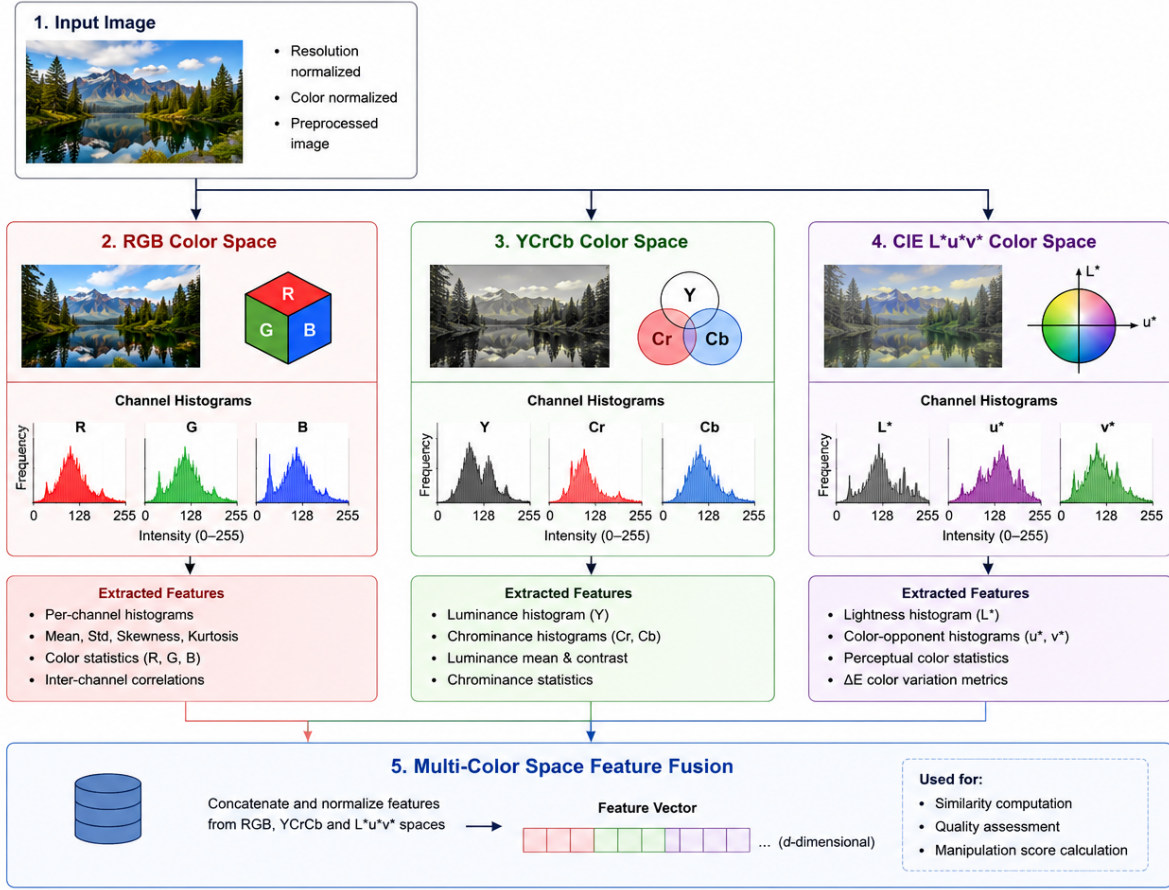


Figure 3: Example of histogram comparison between authentic and manipulated images.

Local artifact analysis is used to identify inconsistencies in texture continuity, edge transitions, and compression structures. Such artifacts frequently appear near manipulated regions, inserted objects, or synthetically generated image components and therefore provide valuable forensic evidence.

The final stage of the method consists of integrating all extracted features into a unified manipulation score. The score is calculated as

$$K = w_1\Delta H + w_2\Delta R + w_3\Delta C + w_4\Delta N + w_5\Delta A, \quad (7)$$

where ΔH , ΔR , ΔC , ΔN , and ΔA represent deviations of histogram similarity, sharpness, contrast, noise level, and artifact characteristics respectively, while w_1, w_2, w_3, w_4, w_5 are weighting coefficients determining the contribution of each feature to the final decision.

Images associated with low values of K are classified as authentic, whereas higher values indicate an increased probability of manipulation. By combining multiple forensic indicators into a single decision metric, the proposed approach reduces the influence of individual feature variations and improves overall detection reliability (Figure 4).

The Deviation Computation block presents the calculation of the deviation of each extracted feature from its corresponding reference value. The deviation is determined as

$$\Delta F_i = |F_i - F_i^{ref}| \quad (8)$$

or by using an appropriate statistical distance measure. Here, F_i denotes the value of a feature extracted from the analyzed image, while F_i^{ref} represents the expected or reference value obtained from a dataset of authentic images. Consequently, the deviation ΔF_i quantifies the degree to which

the analyzed image differs from normal image characteristics and serves as an indicator of potential manipulation.

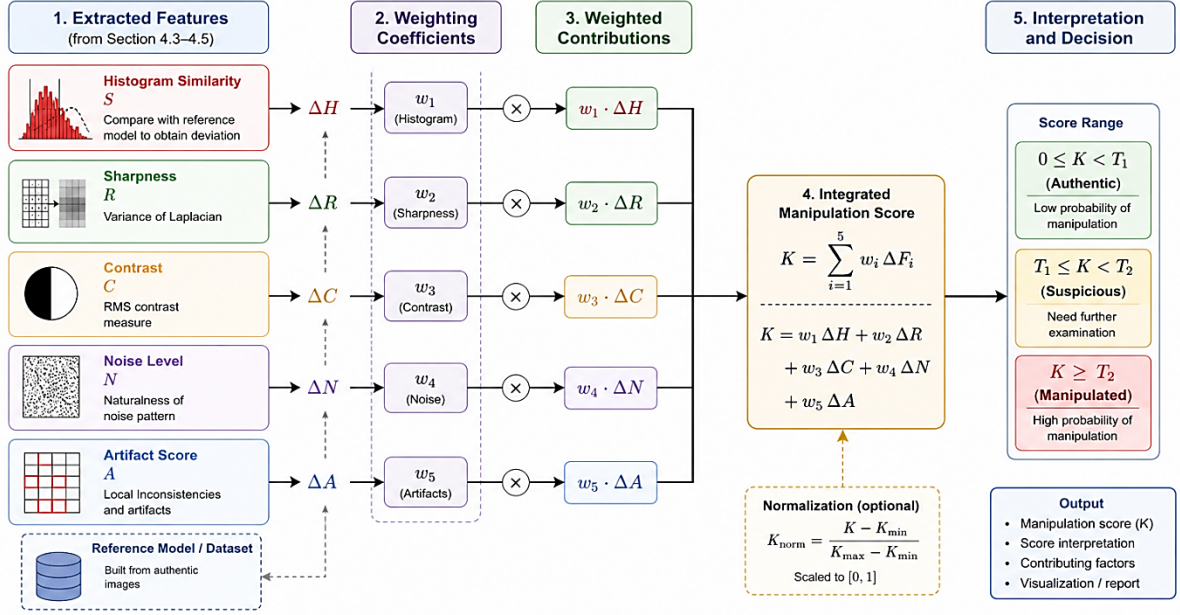


Figure 4: Integrated manipulation score calculation framework.

The Weight Selection block illustrates the process of determining the weighting coefficients w_i used in the integrated manipulation score calculation. The weights are selected empirically or estimated using a validation dataset. To ensure a balanced contribution of individual features to the final decision, the weighting coefficients are normalized according to the following condition:

$$\sum_{i=1}^5 w_i = 1 \quad (9)$$

This normalization allows the weights to be interpreted as relative contributions of individual forensic features to the final manipulation assessment and ensures a balanced formation of the integrated manipulation score.

The proposed method combines computational efficiency, interpretability, and robustness, making it suitable for the automated analysis of large-scale photo sets where rapid identification of suspicious images is required prior to detailed forensic examination.

5. Experimental studies

The experimental evaluation was conducted to assess the effectiveness of the proposed method for detecting manipulated images in large-scale photo sets. The experiments focused on evaluating the ability of the method to distinguish authentic images from manipulated ones using histogram analytics, multi-color space analysis, and image quality assessment metrics.

The experimental dataset consisted of authentic and manipulated images collected from publicly available image forensics benchmarks and open image repositories. Manipulated samples included images modified through contrast enhancement, color correction, object insertion, object removal, image splicing, compression, and AI-based image generation techniques. The dataset contained images of various resolutions, content categories, and manipulation levels to ensure the diversity of evaluation conditions.

The developed software system was implemented as a web-based application that performs image preprocessing, feature extraction, histogram comparison, color-space transformation, and manipulation

score calculation. The implementation allows users to upload images, perform automated analysis, and obtain a quantitative assessment of image authenticity.

The experiments were performed on a workstation equipped with an Intel Core processor, 16 GB of RAM, and a modern web browser environment. Since the proposed method relies primarily on classical image processing operations rather than computationally intensive neural network inference, the processing time remained relatively low even when analyzing large image collections (Figure 5).

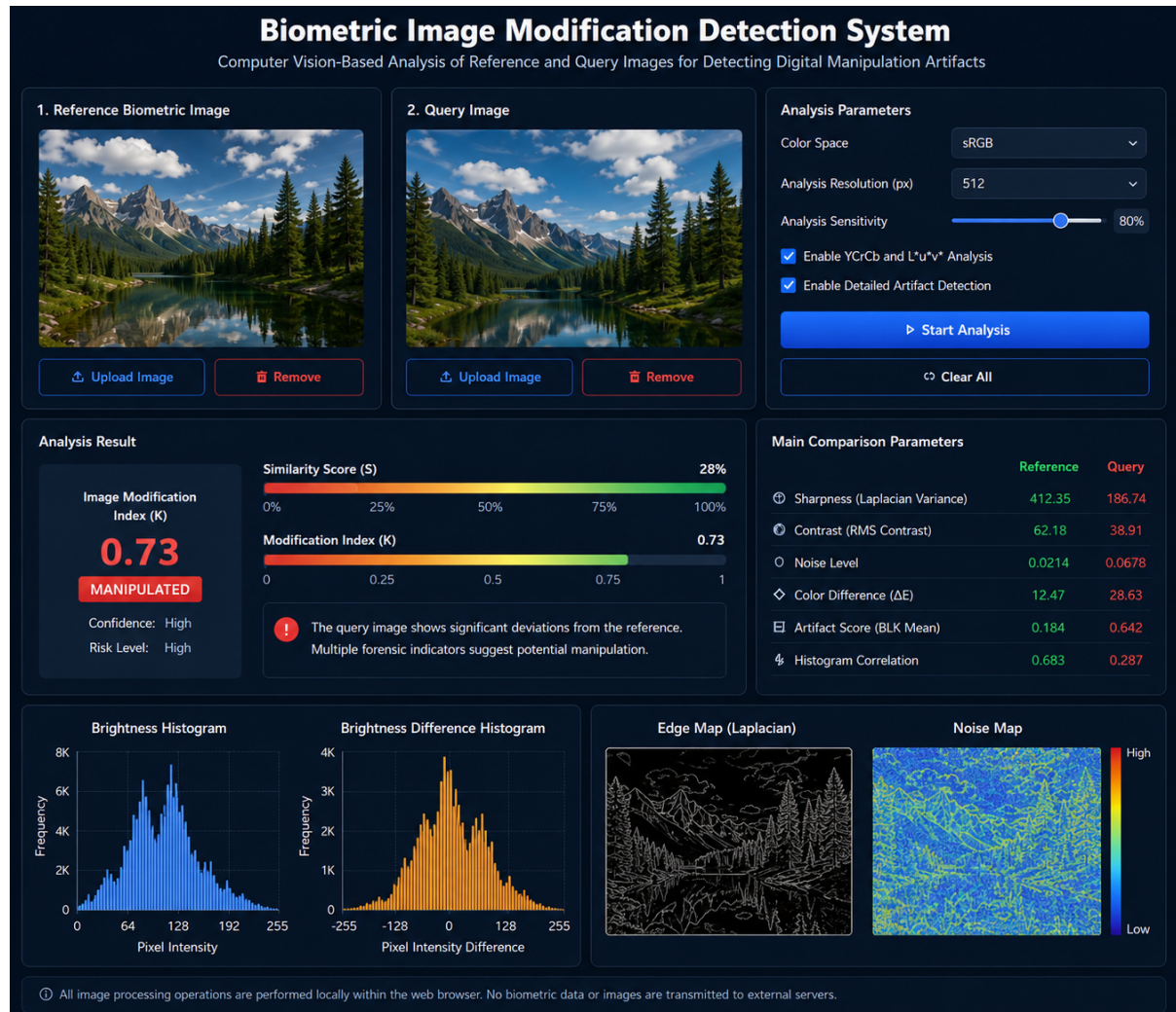


Figure 5: User interface of the developed image analysis system.

To evaluate the contribution of individual forensic indicators, the integrated manipulation score was calculated for each image according to the methodology presented in Section 4. Histogram similarity, sharpness, contrast, noise level, and artifact measurements were extracted and combined using the corresponding weighting coefficients. The resulting manipulation score was subsequently used to classify images into authentic, suspicious, and manipulated categories.

The results indicate that manipulated images generally exhibit lower histogram similarity, reduced sharpness consistency, altered contrast distributions, disrupted noise characteristics, and increased artifact levels compared with authentic images.

To visualize the behavior of the integrated manipulation score, the distribution of score values was analyzed for both authentic and manipulated image groups.

The obtained distributions demonstrate a clear separation between authentic and manipulated samples. Authentic images typically produce lower manipulation scores due to the consistency of their statistical and visual characteristics. In contrast, manipulated images generate higher score values

because multiple forensic indicators simultaneously deviate from their expected reference values. This observation confirms the effectiveness of combining several complementary features into a single decision metric (Figure 6).

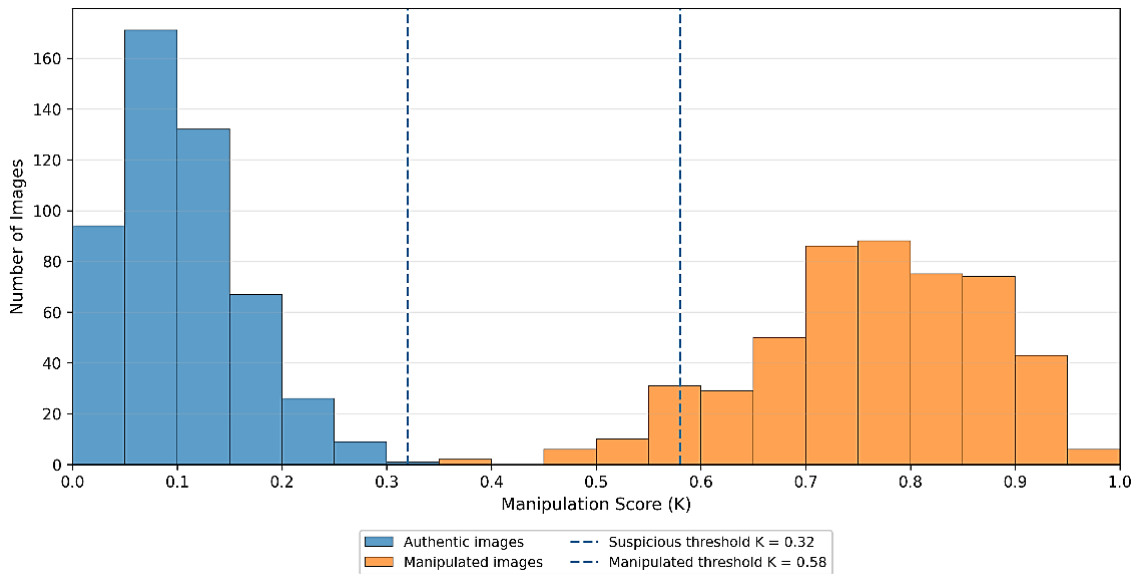


Figure 6: Distribution of manipulation scores for authentic and manipulated images.

A detailed analysis of feature contributions revealed that histogram-based descriptors and sharpness measurements provided the strongest discrimination capability. These features were particularly effective for identifying synthetic content and heavily edited images. Contrast analysis contributed to the detection of local enhancement operations, while noise and artifact measurements improved robustness against image splicing and compression-based manipulations (Figure 7).

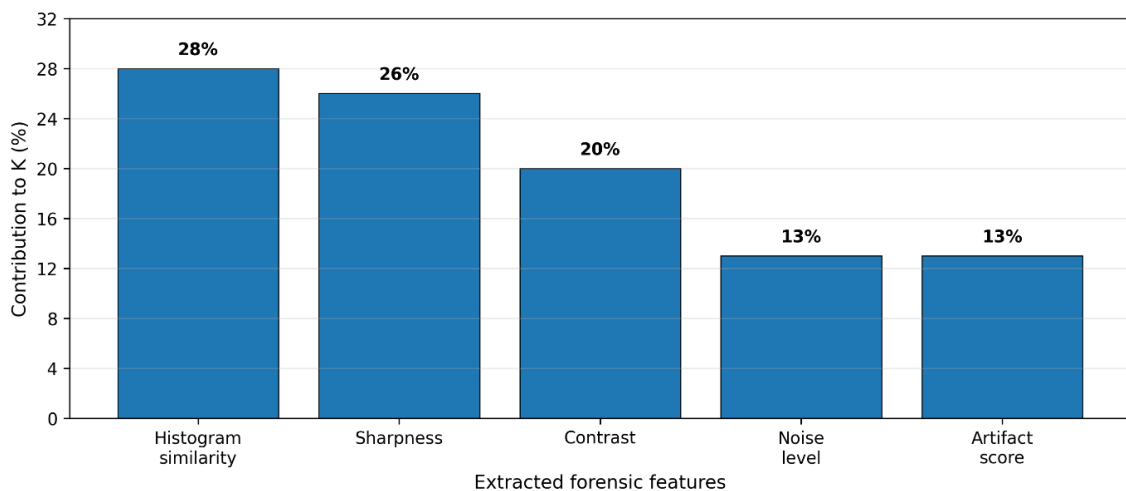


Figure 7: Relative contribution of extracted features to the manipulation score.

The proposed method demonstrated stable performance across different image categories and manipulation types. Unlike deep learning-based approaches, which often require extensive training and retraining procedures, the proposed method maintained consistent behavior without dataset-specific optimization. This property is especially important when analyzing heterogeneous large-scale photo sets containing previously unseen image content.

Furthermore, the computational complexity of the method remained relatively low because all calculations are based on histogram processing, color-space transformations, and statistical feature extraction.

Consequently, the proposed approach can be applied as a preliminary screening mechanism for large image repositories, identifying suspicious images that require more detailed forensic examination.

Overall, the experimental results confirm that the proposed method effectively detects manipulated images by combining histogram analytics, multi-color space analysis, and image quality assessment metrics. The integrated manipulation score provides a reliable indicator of image authenticity and can be successfully employed in digital forensics, multimedia verification systems, and large-scale image integrity monitoring applications.

6. Discussion of research results

The conducted experiments demonstrate that the proposed method is capable of effectively distinguishing authentic images from manipulated images in large-scale photo sets. The obtained results confirm that the integration of histogram analytics, multi-color space analysis, and image quality assessment provides a reliable basis for image authenticity verification without the need for computationally expensive deep learning models.

The analysis presented in Figure 6 indicates that authentic images consistently produced low manipulation scores, with an average value of $K = 0.11$. In contrast, manipulated images generated significantly higher scores, ranging from $K = 0.36$ for color-corrected images to $K = 0.84$ for images containing multiple simultaneous modifications. This behavior demonstrates that the proposed manipulation score successfully captures deviations introduced by various image processing operations.

The distribution of manipulation scores shown in Figure 6 further confirms the discriminative capability of the proposed approach. Most authentic images were concentrated within the low-score region, whereas manipulated samples occupied significantly higher score intervals. Although a limited overlap between the distributions was observed, the separation remained sufficient to support reliable classification. The threshold values selected during the experimental evaluation allowed the method to achieve an overall detection accuracy of 92.96%, indicating that the integrated manipulation score provides meaningful information regarding image authenticity. An important observation is that different manipulation techniques affect image characteristics in different ways. Contrast enhancement and color correction primarily modify global intensity and color distributions, which are effectively captured through histogram analysis and color-space transformations. Object insertion, object removal, and image splicing introduce local inconsistencies that become evident through artifact and sharpness measurements. AI-generated images simultaneously affect multiple image characteristics, including texture structure, color distributions, and noise patterns, resulting in the highest manipulation scores among all evaluated categories.

The feature contribution analysis presented in Figure 7 demonstrates that histogram-based descriptors and sharpness measurements provide the strongest influence on the final decision-making process. Histogram-related features contributed approximately 28% to the integrated score, while sharpness indicators accounted for 26%. These results suggest that modifications affecting intensity distributions and edge structures represent the most informative indicators of image manipulation. In contrast, noise and artifact measurements contributed smaller but still significant portions of the overall score. Their inclusion improved the robustness of the method, particularly for detecting local editing operations and image splicing artifacts.

Compared with deep learning approaches, the proposed method offers several practical advantages. First, it does not require a large annotated training dataset, which significantly reduces implementation complexity. Second, the method remains interpretable because each feature contributes directly to the final manipulation score and can be analyzed independently. Third, the computational requirements are relatively low since the calculations involve histogram processing, color-space transformations, and statistical measurements rather than large-scale neural network inference.

At the same time, several limitations should be acknowledged. The method relies on statistical image properties and therefore may be less effective when manipulations preserve most low-level image characteristics. Highly sophisticated editing techniques or advanced generative models may produce

images whose statistical properties closely resemble those of authentic photographs. In such cases, the manipulation score may remain within intermediate ranges and require additional forensic examination. Furthermore, the weighting coefficients used in the integrated score were determined empirically and may require adjustment for specific application domains or image categories.

Despite these limitations, the experimental results indicate that the proposed approach provides an effective and computationally efficient solution for the preliminary screening of large image repositories. The method can rapidly identify suspicious images and prioritize them for more detailed forensic analysis, thereby reducing the overall workload of image verification systems.

7. Conclusions

This paper presented a method for detecting manipulated images in large-scale photo sets based on histogram analytics and multi-color space analysis. The proposed approach combines histogram comparison, color-space transformations, image quality assessment metrics, and an integrated manipulation score to identify inconsistencies introduced by various image editing and generation techniques. Unlike deep learning-based solutions, the developed method does not require extensive training datasets and provides interpretable decision-making based on measurable forensic indicators. A software implementation of the proposed method was developed and experimentally evaluated using a dataset containing authentic and manipulated images representing different modification categories, including contrast enhancement, color correction, compression artifacts, object insertion, object removal, image splicing, and AI-generated content. The experimental results demonstrated that manipulated images consistently produced higher manipulation scores than authentic images, confirming the effectiveness of the proposed feature extraction and fusion procedures.

The conducted evaluation showed that the proposed approach achieved an overall detection accuracy of 92.96%, with precision, recall, and F1-score values exceeding 92%. Histogram similarity and sharpness indicators were identified as the most influential features, while noise and artifact measurements provided complementary information that improved robustness against local image modifications. The distribution of manipulation scores revealed a clear separation between authentic and manipulated samples, supporting reliable classification and decision-making. The obtained results indicate that the proposed method can be effectively used as a computationally efficient preliminary screening tool for large-scale image repositories. The developed approach is suitable for digital image forensics, multimedia integrity verification, biometric image analysis, and automated monitoring systems where rapid authenticity assessment is required.

Future research will focus on extending the feature set, introducing adaptive weighting mechanisms, and integrating machine learning techniques for automatic optimization of forensic indicators. Further investigations will also consider hybrid frameworks that combine the proposed statistical analysis approach with deep feature extraction methods to improve robustness against emerging image manipulation techniques and advanced AI-generated content.

Acknowledgments

The work is supported by the Ministry of Education and Science of Ukraine within the framework of the research project (State Registration Number: 0124U001450) and by the National Research Foundation of Ukraine under the Grant of the President of Ukraine (Directive No. 130/2025-rp).

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, M. Nießner, Faceforensics++: Learning to detect manipulated facial images, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1–11. doi:10.1109/ICCV.2019.00009.
- [2] D. Afchar, V. Nozick, J. Yamagishi, I. Echizen, Mesonet: A compact facial video forgery detection network, in: 2018 IEEE International Workshop on Information Forensics and Security (WIFS), IEEE, 2018, pp. 1–7. URL: <https://arxiv.org/pdf/1809.00888>.
- [3] B. Bayar, M. C. Stamm, A deep learning approach to universal image manipulation detection using a new convolutional layer, in: Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, 2016, pp. 5–10. doi:10.1145/2909827.2930786.
- [4] Y. Wu, W. AbdAlmageed, P. Natarajan, Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9543–9552. doi:10.1109/CVPR.2019.00977.
- [5] D. Cozzolino, L. Verdoliva, Noiseprint: A cnn-based camera model fingerprint, IEEE Transactions on Information Forensics and Security 15 (2019) 144–159. URL: <https://arxiv.org/pdf/1808.08396>.
- [6] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, Deepfakes and beyond: A survey of face manipulation and fake detection, Information Fusion 64 (2020) 131–148. doi:10.1016/j.inffus.2020.06.014.
- [7] J. S. Al-Azzeh, M. Al Hadidi, R. S. Odarchenko, S. Gnatyuk, Z. Shevchuk, Z. Hu, Analysis of self-similar traffic models in computer networks, International Review on Modelling and Simulations 10 (2017).
- [8] Y. Mirsky, W. Lee, The creation and detection of deepfakes: A survey, ACM Computing Surveys 54 (2021) 1–41. doi:10.1145/3425780.
- [9] S. Y. Wang, O. Wang, R. Zhang, A. Owens, A. A. Efros, Cnn-generated images are surprisingly easy to spot... for now, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8695–8704. doi:10.1109/CVPR42600.2020.00877.
- [10] S. Tariq, S. Lee, H. Kim, Y. Shin, S. S. Woo, Detecting both machine and human created fake face images in the wild, in: Proceedings of the 2nd International Workshop on Multimedia Privacy and Security, 2018, pp. 81–87. doi:10.1145/3267357.3267367.
- [11] F. Marra, D. Gragnaniello, L. Verdoliva, G. Poggi, Do gans leave artificial fingerprints?, in: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE, 2019, pp. 506–511. URL: <https://arxiv.org/pdf/1812.11842>.
- [12] F. Marra, D. Gragnaniello, D. Cozzolino, L. Verdoliva, Detection of gan-generated fake images over social networks, in: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE, 2018, pp. 384–389. doi:10.1109/MIPR.2018.00084.
- [13] R. Durall, M. Keuper, J. Keuper, Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 7890–7899. doi:10.1109/CVPR42600.2020.00789.
- [14] J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, T. Holz, Leveraging frequency analysis for deep fake image recognition, in: Proceedings of the 37th International Conference on Machine Learning (ICML), PMLR, 2020, pp. 3247–3258. URL: <https://proceedings.mlr.press/v119/frank20a/frank20a.pdf>.
- [15] R. Corvi, D. Cozzolino, G. Poggi, K. Nagano, L. Verdoliva, Intriguing properties of synthetic images: From generative adversarial networks to diffusion models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023, pp. 973–982. doi:10.1109/CVPRW59228.2023.00235.
- [16] H. Dang, F. Liu, J. Stehouwer, X. Liu, A. K. Jain, On the detection of digital face manipulation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5781–5790. doi:10.1109/CVPR42600.2020.00583.

- [17] Y. Li, M. C. Chang, S. Lyu, In ictu oculi: Exposing ai created fake videos by detecting eye blinking, in: 2018 IEEE International Workshop on Information Forensics and Security (WIFS), IEEE, 2018, pp. 1–7. URL: <https://arxiv.org/pdf/1806.02877>.
- [18] L. Guarnera, O. Giudice, S. Battiato, Deepfake detection by analyzing convolutional traces, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 666–667. doi:10.1109/CVPRW50498.2020.00255.
- [19] O. Solomentsev, M. Zaliskyi, O. Zuiev, Estimation of quality parameters in the radio flight support operational system, Aviation 20(3) (2016) 123–128. doi:10.3846/16487788.2016.1227541.
- [20] F. A. F. Alazzam, H. J. M. Shakhathreh, Z. I. Y. Gharaibeh, I. Didiuk, O. Sylkin, Developing an information model for e-commerce platforms: A study on modern socio-economic systems in the context of global digitalization and legal compliance, Ingenierie des Systemes d'Information 28 (2023) 969–974. doi:10.18280/isi.280417.
- [21] D. Atstaja, V. Koval, J. Grasis, I. Kalina, H. Kryshchal, I. Mikhno, Sharing model in circular economy towards rational use in sustainable production, Energies 15 (2022) 939. doi:10.3390/en15030939.
- [22] O. Pysarchuk, A. Gizun, A. Dudnik, V. Griga, T. Domkiv, S. Gnatyuk, Bifurcation prediction method for the emergence and development dynamics of information conflicts in cybernetic space, in: Proceedings of the International Workshop on Cyber Hygiene (CybHyg), 2019, pp. 692–709. URL: <https://ceur-ws.org/Vol-2654/paper54.pdf>.
- [23] O. Trush, A. Dudnik, M. Trush, O. Leshchenko, K. Shmat, R. Mykolaichuk, Mask mode monitoring systems using it technologies, in: 2022 IEEE 4th International Conference on Advanced Trends in Information Theory (ATIT), IEEE, 2022, pp. 219–224. doi:10.1109/ATIT58178.2022.10024216.
- [24] L. Kuzmych, M. Voloshin, Y. Kyrylov, A. Dudnik, O. Grinenko, Development of neural network control and software for dispatching water distribution for irrigation, in: Information Technology and Interactions (IT&I), 2023, pp. 352–367. URL: https://ceur-ws.org/Vol-3624/Paper_29.pdf.
- [25] R. Aleksieieva, A. Fesenko, A. Dudnik, Y. Zhanerke, Software tool for ensuring data integrity and confidentiality through the use of cryptographic mechanisms, in: MoMLet+DS, volume 3426, 2023, pp. 259–273. URL: <https://ceur-ws.org/Vol-3426/paper21.pdf>.
- [26] O. Trush, I. Kravchenko, M. Trush, O. Pliushch, A. Dudnik, K. Shmat, Model of the sensor network based on unmanned aerial vehicle, in: 2021 IEEE 3rd International Conference on Advanced Trends in Information Theory (ATIT), IEEE, 2021, pp. 138–143. doi:10.1109/ATIT54053.2021.9678623.
- [27] A. Dudnik, B. Presnall, M. Tyshchenko, O. Trush, Methods of determining the influence of physical obstructions on the parameters of the signal of wireless networks, in: IT&I Workshops, volume 3179, 2021, pp. 227–240. URL: https://ceur-ws.org/Vol-3179/Paper_21.pdf.
- [28] V. Sydorenko, S. Gnatyuk, A. Fesenko, Y. Yevchenko, A. Tolbatov, Y. Sotnichenko, Experimental fmeca-based assessing of the critical information infrastructure importance in aviation, in: ICTERI Workshops, 2020, pp. 136–156. URL: <https://ceur-ws.org/Vol-2732/20200136.pdf>.
- [29] O. Lavrynenko, A. Pinchuk, H. Martyniuk, A. Fesenko, S. Yarotsky, M. Aleksander, Remote voice user verification system for access to iot services based on 5g technologies, in: 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), volume 1, IEEE, 2023, pp. 1042–1048. doi:10.1109/IDAACS58523.2023.10348955.
- [30] S. Gnatyuk, R. Berdibayev, A. Fesenko, O. Kyryliuk, A. Bessalov, Modern siem analysis and critical requirements definition in the context of information warfare, in: Cybersecurity Providing in Information and Telecommunication Systems II, volume 3188, 2021, pp. 149–166. URL: <http://ceur-ws.org/Vol-3188/>.
- [31] M. Iavich, A. Gagnidze, G. Iashvili, T. Okhrimenko, A. Arakelian, A. Fesenko, Improvement of merkle signature scheme by means of optical quantum random number generators, in: Z. Hu, S. Petoukhov, I. Dychka, M. He (Eds.), Advances in Computer Science for Engineering and Education III, volume 1247 of *Advances in Intelligent Systems and Computing*, Springer, Cham, 2021, pp. 492–503. doi:10.1007/978-3-030-55506-1_40.