Detecting AI-Generated Synthetic Images using Multichannel Forensic Computer Vision and CNN Ensembles

Tom Brammer, Andreas Iakeimiadis, Niklas Minkowitsch, Rune Molzen, Fabian Thies, Laura Witt

Hochschule Flensburg
Flensburg, Schleswig-Holstein, Germany

Figure 1: Generated image representation of the fictitious arrest of Donald Trump (Midjourney). Tweet by Eliot Higgins (@EliotHiggins) on X [1] .

Figure 2: Instagram post by the AfD, diagnosed as AI-generated (Hubert, Fraunhofer Institute Stuttgart, 2024). The image quality is assessed as low, presumably due to an outdated AI model or low effort [2, 3].

Abstract

Artificial Intelligence-generated synthetic images pose an increasing threat to digital security, media integrity, and trust in information systems. The project aims to combine classical image processing methods with modern deep learning techniques to enable reliable identification of image manipulations. The ensemble model developed is based on the EfficientNetV2 architecture and combines forensic computer vision methods such as Fast Fourier Transform (FFT) and Spatial Rich Model (SRM) features to reliably detect synthetic patterns of such images.

Training was conducted using a multichannel approach, ensemble strategy, and train-time augmentation, which sustainably strengthened the robustness and generalizability of the models. The initial results of the evaluation show a promising accuracy of 94% with a selected test set including images based on state-of-the-art generative image models. Thus, the work provides a solid foundation for further research in the field of digital image forensics.

1 Introduction

In today's digital era, images and videos play an increasingly central role as sources of information. At the same time, manipulated media created using generative and deepfake technologies are spreading rapidly, and are deliberately employed for the dissemination of false information, identity fraud, and other criminal purposes [4,5]. For example, between July 2023 and July 2024, a total of 82 politically motivated deepfakes were identified across 38 countries[6]. These deepfakes targeted public figures such as heads of state, candidates, and journalists with the intent to influence elections, conduct character assassination, or spread disinformation[5,6]. Notable cases include manipulative videos of politicians like the Canadian Prime Minister Justin Trudeau and the Turkish President Erdoğan, where deepfakes were used to falsely associate an opposition leader with terrorist groups [5,7]. Additionally, deepfake technologies were employed during the 2024 US election campaign, including fake phone calls and manipulated videos that undermined voter trust and endangered political processes[8]. These developments highlight the urgency of developing effective detection mechanisms to protect digital security and the integrity of democratic processes.

Against this backdrop, the present work focuses on the development of a convolutional neural network (CNN) based method for the automatic detection of AI-generated synthetic images. The goal is to create a robust and accurate solution by combining classical image analysis techniques with modern deep learning models to reliably identify manipulations even in complex forensic scenarios. Special emphasis is placed on the use of forensic features such as frequency analyses and specific noise patterns to effectively detect subtle traces of manipulation.

This research thus represents an important contribution to addressing the growing challenge posed by synthetic generated media in society and politics.

2 Related Work

Research on the detection of forged images ("fake-image detection") has made significant progress in recent years. Initially, classical approaches such as metadata analysis (Exif) dominated, where image information such as the creation date, camera model, editing history, or GPS data are scrutinized for inconsistencies[9]. Techniques like color and lighting analysis identify discrepancies in light direction, color dynamics, or shadow progression and are established tools in image forensics[10]. However, these methods reach their limits because metadata can be easily manipulated, and visual effects can be deliberately replicated using deepfake technologies.

With the advent of machine learning techniques, image authenticity began to be automatically verified. Earlier ML approaches mainly employed convolutional neural networks (CNNs) based on RGB image data, using models such as VGG, ResNet, and early variants of EfficientNet[11]. These models delivered solid results, particularly for simple forgeries, but were often too specialized and lacked robustness when confronted with generative AI images.

Current research is increasingly based on deep learning models with extended feature channels. Particularly successful are methods that, in addition to RGB data, utilize forensic channels such as Spatial Rich Model (SRM) filters or Fourier transformations (FFT) [12]. SRM filters extract local noise structures and manipulation-specific texture patterns, significantly enhancing detection capabilities. Studies demonstrate that preprocessing with SRM filters and frequency features (e.g., via FFT) can improve detection rates by several percentage points, especially when combined with a modern CNN backbone architecture like EfficientNetV2.

International research initiatives such as FakeBench focus on developing explainable multimodal models for synthetic image detection and attribution [13]. Systems like Smogy or proprietary software from Microsoft also serve as important benchmarks in automated image forensics.

Despite notable progress in forgery detection, significant research gaps remain. Many current models specialize in certain types of image manipulation and demonstrate limited capability to adequately handle previously unseen or out-of-distribution manipulations. Furthermore, a lack of explainability in model decisions complicates their forensic and legal application. The ability of models to generalize to novel AI-generated synthetic images is rarely reliably achieved in practice. Additionally, many approaches neglect multimodal integration, typically focusing solely on RGB-based features.

The scripts developed in this research project, covering both training and inference, specifically address the shortcoming of lacking generalization outside of used training data. The project implements multimodal feature extraction using RGB, Fast Fourier Transform (FFT), and Spatial Rich Model (SRM), combined with modern deep learning architectures, in particular EfficientNetV2 [14]. The objective is to improve robustness and out-of-distribution generalization, enabling the reliable detection of previously unseen and complex forgeries.

3 Methodology

To achieve a robust CNN-based classification of AI-generated synthetic images, a solid dataset is needed to train the difference between real and fake images. While training, the images are preprocessed with further discussed computer vision methods.

3.1 Dataset

For model training and evaluation, a comprehensive and diversified dataset was assembled, comprising 34,100 genuine images (referred to as "Real") and 34,100 artificially generated images (referred to as "Fake"). Fake images that are not trivially distinguishable as non-authentic within short viewing durations were carefully selected to ensure a high quality starting point. Since existing public datasets were either outdated or insufficient, a mixed dataset was created by combining various sources and generating images independently.

The genuine images originate from a dataset provided by the University of Trento and a Real-Image set derived from scientific publications [12, 15]. These were supplemented by social media images obtained from social networks (Reddit, X [formerly Twitter], Telegram, Facebook) and free image repositories (Unsplash, Pixabay). In addition, scraping scripts were employed to ensure the diversity and volume of real image data.

The "Fake" set was generated using modern generative models, including state-of-the-art systems such as Dall-E, ChatGPT, Grok, Flux, Perplexity, StyleGAN V2, Midjourney up to version 6, and Stable Diffusion XL (SDXL) and SD1.5. When compatible, a wide selection of LoRAs was used, aiming to achieve higher levels of photorealism. This wide range of generation techniques guarantees a broad coverage of current image generators and image types.

For data preprocessing, all images were cropped and normalized. The dataset was also cleaned by removing irrelevant or erroneous entries. The preprocessing followed a standardized procedure to ensure a consistent data foundation. The dataset was divided into training and validation subsets using a classic 80:20 ratio, with random seeds applied to ensure fair evaluation and a balanced representation of both classes in the dataset.

It should be noted that a dataset size of 68,200 images is not considered to be large on the subject of training a convolutional neural network as it is done in this approach. However, since the chosen images are all high quality images that represent the type of synthetic images that could be mistaken for real, and it is aimed to solve a binary classification problem, it suffices as a solid basis.

3.2 Detection Model Architecture

3.2.1 Architecture Foundation. The proposed approach employs an ensemble of EfficientNetV2 architectures, selected for their exceptional balance between classification accuracy and computational efficiency [14, 17, 18]. The fundamental innovation lies in the integration of multi-channel forensic analysis with contemporary deep learning methodologies, enabling robust detection of AI-generated images across diverse generation paradigms .

3.2.2 Multi-Channel Input Design. The standard three-channel RGB input architecture is extended to accommodate nine distinct channels, capturing both natural image characteristics and forensically relevant manipulation signatures. The input structure comprises three RGB channels. Additionally, the architecture incorporates a single Fast Fourier Transform channel representing frequency domain characteristics, and five Spatial Rich Model channels designed to capture high-frequency statistical anomalies indicative of synthetic generation processes.

3.2.3 Forensic Channel Processing. The FFT component transforms grayscale image data to the frequency domain through a systematic process. The two-dimensional FFT is applied to the grayscale conversion of the input image, followed by frequency shift centering and magnitude computation. This transformation compresses the dynamic range while preserving frequency information critical for synthetic content detection.

The SRM filter bank implementation employs five specialized 3×3 convolution kernels, each designed to detect specific categories of statistical anomalies. These kernels include high-pass directional filters, edge detection variants, and residual noise extractors.

3.2.4 Architectural Adaptations. To accommodate the nine-channel input while leveraging pre-trained ImageNet weights, two architectural configurations are supported. The direct mode integrates nine channels natively into the model architecture. The stem adapter mode implements a learned projection layer consisting of a 1×1 convolution mapping nine channels to three, followed by batch normalization and SiLU activation. This configuration enables utilization of pre-trained backbone weights while adapting to forensic inputs.

3.2.5 Model Selection and Ensemble Configuration. Initial experiments encompassed training EfficientNetV2 variants across the complete architecture spectrum, including B0, B1, B2, B3, S, M, L, and XL configurations. Performance analysis revealed that optimal parameter ranges exist within the B2, B3, S, and M variants. The B0 and B1 models demonstrated insufficient capacity, exhibiting poor performance due to limited parameter counts. Conversely, the L and XL variants achieved performance comparable to the M model but showed signs of overfitting, indicating excessive capacity for the available training data.

3.2.6 Ensemble Composition. Through systematic evaluation of 30 individual models with varying architectures, random seeds, and train-validation split ratios (ranging from 0.18 to 0.21), the final ensemble composition was determined. The optimal configuration consists of five models: one B2, one B3, and three S variants, each trained with different random seeds and slightly varied train-validation ratios. The lower train-validation ratios (e.g., 0.2, corresponding to 20% validation and 80% training data) consistently produced superior results, suggesting that further dataset expansion could yield additional accuracy improvements. The models utilized in the ensemble use confidence-weighed scores to determine the weight each models individual prediction gets.

3.2.7 Seed Diversification Strategy. Each model variant employs unique random seed offsets to minimize correlation between ensemble members and reduce overfitting to specific data partitions. This approach enhances ensemble diversity and improves robustness across varied input conditions.

3.2.8 Training Protocol. The training protocol implements a two-phase approach to optimize convergence and prevent overfitting. Phase one constitutes a three-epoch warm-up period where only classification head parameters remain trainable. This phase facilitates forensic feature adaptation without disrupting pre-trained representations. Phase two encompasses a ten-epoch full fine-tuning period with all parameters trainable and a reduced learning rate, enabling end-to-end optimization with gradient checkpointing for memory efficiency.

3.2.9 Advanced Optimization Techniques. The optimization framework employs AdamW with weight decay of 0.05 for regularization. Gradient clipping with a maximum norm of 1.0 ensures training stability.

3.2.10 Test-Time Augmentation Framework. The TTA implementation incorporates eight distinct transformations to enhance prediction robustness. These transformations include center crop configurations, horizontal flipping, multi-scale random crops with scale factor 1.1, and light rotations of ±5 degrees. Each transformation preserves geometric consistency while introducing controlled variability for robust inference. TTA predictions undergo averaging across all augmentation variants. This approach reduces prediction variance and enhances reliability for challenging or ambiguous cases.

3.2.11 Threshold Optimization and Manual Adjustment. Individual model decision thresholds initially undergo optimization through precision-recall curve analysis, with F1-score maximization serving as the objective function. Subsequently, manual threshold adjustment was performed through systematic testing of different threshold combinations across ensemble components. This manual fine-tuning process optimizes inter-model cooperation by balancing individual model sensitivities and specificities. The rationale for this approach lies in the heterogeneous nature of ensemble members: different architectures exhibit varying confidence distributions and decision boundaries. By adjusting thresholds to harmonize these differences, the ensemble achieves improved collective performance that exceeds the sum of individual contributions.

3.2.12 Data Augmentation Strategy. Training augmentations include random resized cropping with scale ranges of 0.9-1.0 and aspect ratio constraints of 0.9-1.1, geometric transformations including horizontal flipping and perspective distortion, and quality degradations such as JPEG compression with quality ranges of 60-100% and blur variants. These augmentations enhance model robustness against common post-processing operations encountered in real-world scenarios.

This comprehensive framework combines established forensic analysis principles with state-of-the-art deep learning architectures, creating a robust system capable of detecting sophisticated AI-generated content while maintaining practical deployability across diverse operational environments.

4 Evaluation

The final evaluation was conducted on a balanced test set consisting of 300 real and 300 fake images. This test set was compiled from diverse sources to ensure broad variety. It includes modern generators, among them some that were not part of the training data (e.g., the "nano banana" generator released by Google September 2025), as well as older and classical fakes. This approach guarantees a realistic and practical assessment of the model's generalization capabilities and detection performance.

4.1 Results

The model's performance can be measured with the following metrics: Accuracy, Precision, Recall and F1-Score. The following metrics provide a nuanced evaluation of the model's quality, reflecting both its ability to correctly classify images and maintain a balance between false positives and false negatives.

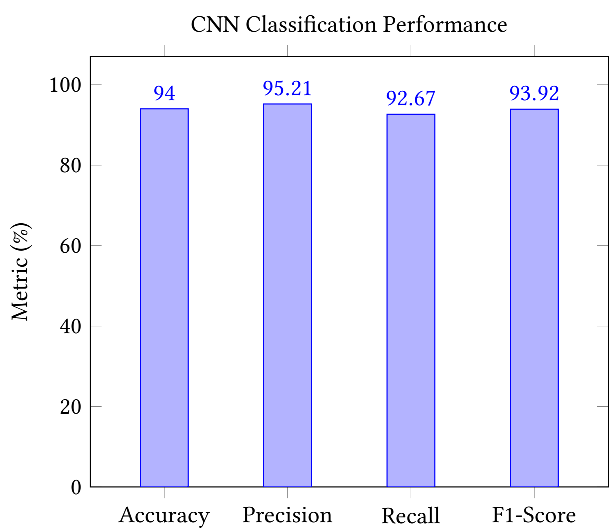

Figure 3: The metrics suggest a balanced hit/miss ratio for True and Fake images.

4.1.1 Confusion Matrix. The confusion matrix illustrates that the classifier achieves a very high hit rate for both classes, with comparatively few misclassifications. A bit more real images are falsely classified as fake than fake images falsely classified as real images.

	Predicted Real	Predicted Fake
Actual Real	286	22
Actual Fake	14	278

Figure 4: The confusion matrix shows out of 300 real and 300 fake images, 22 real and 14 fake images were misclassified.

4.1.2 Single Model vs. Ensemble. While individual EfficientNetV2 models already demonstrate solid results with accuracies between 89.9% and 92.5%, the ensemble strategy delivers the best performance in terms of accuracy and robustness. Combining multiple models trained with different seeds and test-validation configurations reduces misclassifications, especially for challenging or uncommon deepfake variants and realistic images.

4.1.3 Context of Results. The metrics illustrate the system's ability to reliably distinguish between currently common AI-generated synthetic images and real images, highlighting the high practical relevance of the method. Notably, the correct detection of synthetic images produced by models not represented in the dataset (such as Googles "nano banana" released 09.2025) evidences the CNN ensemble's broad generalization capability.

5 Comparison with baseline methods

Traditionally, the detection of forged images primarily relied on classical computer vision methods. A widely used approach is the analysis of image information encoded in EXIF metadata, which provides details about the capture device, capture date, and any editing steps performed. These methods are complemented by color and lighting analyses aimed at uncovering inconsistencies in light direction, color dynamics, or shadow patterns.

5.1 Results of the manual method

To assess the performance of classical methods, a manually created baseline pipeline was evaluated. The results clearly demonstrate the limited practical applicability of such approaches:

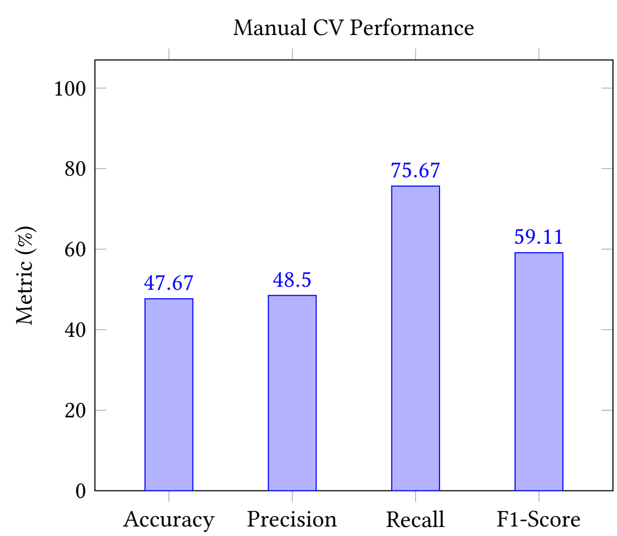

Figure 5: Metrics of manual weighed CV classification shows unsatisfying results with the same diversified, modern testset.

These metrics indicate that classical approaches lag significantly behind modern deep learning models. While the relatively high recall shows that many fake images are detected, a large number of genuine, unaltered images are mistakenly classified as manipulated, which severely restricts their usability in practice.

	Predicted Real	Predicted Fake
Actual Real	59	73
Actual Fake	241	227

Figure 6: The confusion matrix for manual weighed CV detection again demonstrates the high amount of falsely classified images.

5.1.1 *Causes of the weaknesses of classical methods.*
A primary cause of the limited effectiveness of rule-based computer vision methods is image post-processing, including filters (e.g., black-and-white effects) and enhancement operations that alter color distributions and contrast, thereby obscuring forensic cues and making manipulated images resemble conventionally edited photographs. Additionally, optical characteristics such as depth-of-field blur attenuate edge sharpness and degrade the reliability of edge-centric analyses. These variations severely affect approaches based on color balance, lighting consistency, or edge detection, which depend on stable, hand-crafted cues. In contrast, a CNN ensemble with multimodal inputs (RGB, FFT, SRM) learns more invariant feature representations and is less susceptible to such transformations, improving robustness under heterogeneous post-processing. Moreover, manual pipelines frequently require per-generator configuration and iterative threshold tuning to attain acceptable detection rates, resulting in substantial maintenance overhead and limiting scalability in practical deployments.

5.2 Comparison with Classical Methods

Classical techniques, including metadata forensics and color or lighting analyses, typically report accuracies in the range of 70%–80% on manually edited images and legacy manipulation methods, yet they face marked limitations with contemporary AI-generated content due to easily falsified or missing metadata and the diminished stability of artifact patterns. The EfficientNetV2 ensemble with multimodal forensic channels (SRM, FFT) surpasses these baselines by approximately 10–20 percentage points in accuracy and attains an F1-score near 94%, while demonstrating robust generalization to previously unseen manipulation types and maintaining a low false-positive rate.

6 Discussion

The experimental results demonstrate that the proposed multichannel EfficientNetV2 ensemble achieves substantial improvements over classical detection methods, with an accuracy of 94% compared to 47.67% for traditional computer vision approaches. This significant performance gap underscores the effectiveness of combining deep learning architectures with forensic feature channels.

6.1 Analysis of Results

The high precision (95.21%) and recall (92.67%) values indicate that the ensemble effectively balances between detecting synthetic images while minimizing false positives on genuine content. The confusion matrix reveals a slight bias toward classifying real images as fake (22 misclassifications) compared to fake images classified as real (14 misclassifications). This conservative approach is preferable in forensic applications where false negatives (undetected manipulations) pose greater risks than false positives.

The ensemble's ability to correctly identify synthetic content from generators not present in the training data, such as Google's "nano banana" generator, demonstrates robust out-of-distribution generalization. This capability addresses one of the primary limitations of existing detection systems and suggests that the multimodal approach captures fundamental characteristics of synthetic generation processes rather than generator-specific artifacts.

6.2 Methodological Contributions

The integration of forensic channels (FFT and SRM) with RGB data provides complementary information that enhances detection robustness. While RGB channels capture visual artifacts, frequency domain analysis through FFT reveals subtle compression patterns and noise characteristics typically introduced by generative models. The SRM filters effectively capture high-frequency statistical anomalies that persist across different post-processing operations.

The ensemble strategy with diversified random seeds and varying train-validation ratios successfully reduces overfitting while improving generalization. The optimal configuration of five models (B2, B3, and three S variants) represents a balance between computational efficiency and detection performance, avoiding both underfitting (B0, B1) and overfitting (L, XL) observed in preliminary experiments.

6.3 Limitations and Challenges

Despite promising results, several limitations require acknowledgment. The dataset size of 68,200 images, while carefully curated, remains relatively small for deep learning standards. The geographic and demographic bias in training data, predominantly from Western social media platforms, may limit generalizability to diverse global contexts. The evaluation on 600 test images, though carefully selected, represents a limited sample of the rapidly evolving landscape of generative models. The performance against adversarial attacks specifically designed to fool detection systems remains untested, representing a critical vulnerability in real-world deployment scenarios. Furthermore, the computational requirements of the ensemble approach may limit practical deployment in resource-constrained environments, and the lack of explainability mechanisms complicates forensic and legal applications where interpretable decisions are essential.

6.4 Future Research Directions

Several avenues warrant investigation to address current limitations and enhance detection capabilities. Domain-adaptive training strategies could improve robustness against evolving manipulation techniques and reduce dependence on specific generator types. Integration of transformer-based architectures with the current CNN ensemble shows potential for handling larger, more diverse datasets. The development of explainable AI components using methods such as SHAP or LIME would enhance interpretability for forensic applications. Real-time optimization for mobile deployment represents another important direction, enabling broader accessibility and immediate detection capabilities. Additionally, adversarial robustness evaluation and defense mechanisms should be prioritized to ensure reliability against sophisticated attack strategies designed to evade detection systems.

7 Conclusion

This research successfully developed and evaluated a robust approach for detecting AI-generated synthetic images through the integration of multimodal forensic analysis with ensemble deep learning architectures. The proposed EfficientNetV2 ensemble, incorporating RGB, FFT, and SRM channels, achieved 94% accuracy on a diverse test set, significantly outperforming classical detection methods by approximately 46 percentage points. Key contributions include the demonstration that forensic feature channels substantially enhance detection capabilities beyond traditional RGB-based approaches, the validation of ensemble strategies for improved generalization across unseen generation methods, and the establishment of a comprehensive evaluation framework incorporating both modern and classical manipulation techniques. The system's ability to correctly identify synthetic content from generators not represented in training data evidences strong generalization capabilities, addressing a critical limitation of existing detection systems.

The conservative detection bias, favoring false positives over false negatives, aligns well with forensic application requirements where undetected manipulations pose significant security risks. However, limitations including dataset size constraints, computational requirements, and the need for explainability mechanisms highlight important areas for future development. The rapidly evolving landscape of generative AI technologies necessitates continuous adaptation and evaluation of detection systems. This work establishes a solid foundation for practical deepfake detection systems while identifying clear pathways for future improvements in robustness, scalability, and interpretability. The methodological framework and empirical insights provide valuable contributions to the digital forensics community and support the ongoing effort to maintain information integrity in an era of increasingly sophisticated synthetic media.

References

[1] Eliot Higgins. *Deepfake AI video showing Vladimir Putin with warts and other oddities*. Mar. 2023. URL: https://x.com/EliotHiggins/status/1637927681734987777.

[2] Swr Aktuell. *KI: AfD Göppingen wirbt mit einem Fake-Gesicht*. Mar. 2024. URL: https://www.swr.de/swraktuell/baden-wuerttemberg/stuttgart/afd-deepfakes-wahlkampf-100.html.

[3] Instagram User. *Instagram Post*. Nov. 2022. URL: https://www.instagram.com/p/C3JHoDWtXHh/?utm_source=ig_web_copy_link&igsh=NXBobXFibms5NzBp.

[4] Keepnet Labs. *Deepfake Statistics and Trends 2025: Growth, Risks, and Future Insights*. Sept. 2025. URL: https://keepnetlabs.com/blog/deepfake-statistics-and-trends.

[5] Surfshark Research Team. *38 countries have faced deepfakes in elections*. Sept. 2025. URL: https://surfshark.com/research/chart/election-related-deepfakes.

[6] Recorded Future Insikt Group. *Targets, Objectives, and Emerging Tactics of Political Deepfakes*. July 2024. URL: https://www.recordedfuture.com/research/targets-objectives-emerging-tactics-political-deepfakes.

[7] AIAAIC. *Deepfake Justin Trudeau endorses Petro-Canada scam*. Nov. 2023. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/deepfake-justin-trudeau-endorses-petro-canada-scam.

[8] NPR. *Deepfakes, Memes, and Artificial Intelligence in Elections*. Dec. 2024. URL: https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections.

[9] Global Fact Checking. *Invisible Clue: How Metadata Analysis Helps Fight Fakes*. 2025. URL: https://globalfactchecking.com/learning_articles/invisible-clue-how-metadata-analysis-helps-fight-fakes/.

[10] Christian Riess and Elli Angelopoulou. “Scene Illumination as an Indicator of Image Manipulation”. In: *Pattern Recognition Letters* (2010). URL: https://faui1-files.cs.fau.de/public/publications/mmsec/2010-Riess-SIA.pdf.

[11] Nalluri Brahma Naidu et al. “Image Forgery Detection using ResNet50”. In: *International Journal for Research in Applied Science and Engineering Technology (IJRASET)* (Mar. 2024). Open access under Creative Commons Attribution License. DOI: 10.22214/ijraset.2024.59317. URL: https://www.ijraset.com/research-paper/image-forgery-detection-using-resnet50.

[12] Peng Zhou et al. “Learning Rich Features for Image Manipulation Detection”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. URL: https://openaccess.thecvf.com/content_cvpr_2018/CameraReady/2813.pdf.

[13] Yixuan Li et al. “FakeBench: Probing Explainable Fake Image Detection via Large Multimodal Models”. In: *arXiv preprint arXiv:2404.13306* (2024). Submitted April 2024, revised September 2024. URL: https://arxiv.org/abs/2404.13306.

[14] Salford University Research Team. *Deepfake Detection: Evaluating the Performance of EfficientNetV2-B2 on Real vs Fake Image Classification*. June 2024. URL: https://salford-repository.worktribe.com/output/4285867/deepfake-detection-evaluating-the-performance-of-efficientnetv2b2-on-real-vs-fake-image-classification.

[15] Stefano Dell'Anna, Andrea Montibeller, and Giulia Boato. “TrueFake: A Real World Case Dataset of Last Generation Fake Images also Shared on Social Networks”. In: *arXiv preprint arXiv:2504.20658* (2025). Submitted April 2025. URL: https://arxiv.org/abs/2504.20658.

[16] Stefano Dell'Anna and Andrea Montibeller. *TrueFake: A Real World Case Dataset of Last Generation Fake Images also Shared on Social Networks - Official Code Repository*. GitHub repository accompanying the IJCNN 2025 paper. 2025. URL: https://github.com/MMLab-unitn/TrueFake-IJCNN25.

[17] Mingxing Tan and Quoc V. Le. “EfficientNetV2: Smaller Models and Faster Training”. In: *arXiv preprint arXiv:2104.00298* (2021). International Conference on Machine Learning (ICML) 2021. URL: https://arxiv.org/abs/2104.00298.

[18] Mingxing Tan and Quoc V. Le. *EfficientNetV2: Smaller Models and Faster Training*. Google Research Publication. 2021. URL: https://research.google/pubs/efficientnetv2-smaller-models-and-faster-training/.