

Managing and verifying image assets at scale

Dimitrios Karageorgiou
Media Analysis, Verification & Retrieval (MeVer) Group
CERTH

EAB-RPC 2023
19th September 2023



artwork generated using DallE-mini



The MediaVerse Vision



- Set up a **decentralized network of intelligent and accessible tools for digital asset management**, allowing barrier-free usage and integration in target media and platforms.
- Allow professionals and laymen alike to express themselves by publishing multimedia content that may be easily shared and licensed
- Empower European stakeholders to enjoy and produce inclusive, diverse, respectful and credible media experiences.



The MAAM Platform

Next-Gen Media Asset Annotation & Management



AI Captioning & Annotation

Two people helping each other up a mountain at sunset.

2 people



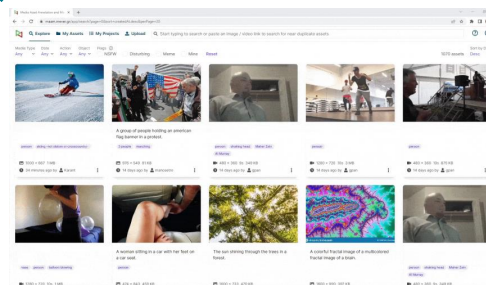
Disturbing Content Protection



A man laying on the ground next to a soldier.

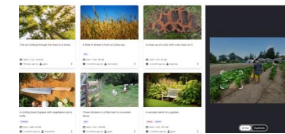


A man laying on the ground next to a soldier.

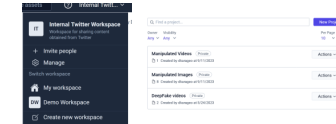


AI-Powered Platform for Managing & Distributing Multimedia Assets

Content-Aware Search



Team Collaboration and Distribution Tools



The MAAM Platform

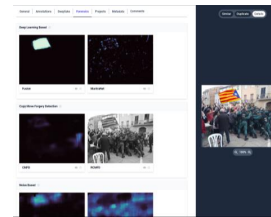
Verifying Untrusted Content

→ In many environments (e.g. journalism) media assets of unknown origin are employed (e.g. breaking news content posted on social media).

→ General-purpose active integrity verification approaches (e.g. digital signatures, blockchain) cannot guarantee the authenticity of an asset before its entry into the trusted environment.

The **MAAM Platform** integrates at its core multiple **Multimedia Verification Tools**

Image Integrity Verification



Deepfake Detection

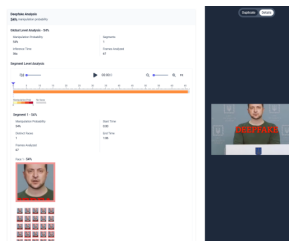


Image Manipulations: Simplicity Wins



→ Despite the recent surge of sophisticated methods for generating fake content, **simple image manipulations remain extremely prevalent.**

(e.g. dominating COVID19-related misinformation, Brennen et al., 2020)

→ Abundance of possible image manipulations.

→ Some manipulations are synonym to uploading to social media platforms: Color enhancements, resizing, cropping etc.

The user should be able to obtain insights regarding different levels of manipulation.



Original



Manipulated

<https://www.businessinsider.com/fake-viral-photo-trump-putin-g20-2017-7>

Image Forgery

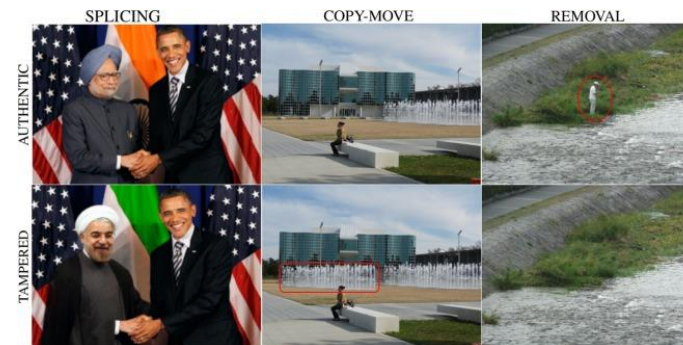


Image manipulations that produce fake graphic content to falsify some facts. (Zheng et. al, 2019)

→ In most scenarios considered as harmful manipulations.

→ **Common cases of image forgery:**

- **Splicing:** Copy-paste a region from a different image.
- **Copy-Move:** Copy-paste a region from the same image.
- **Inpainting:** Fill-in parts of an image using software-generated content.



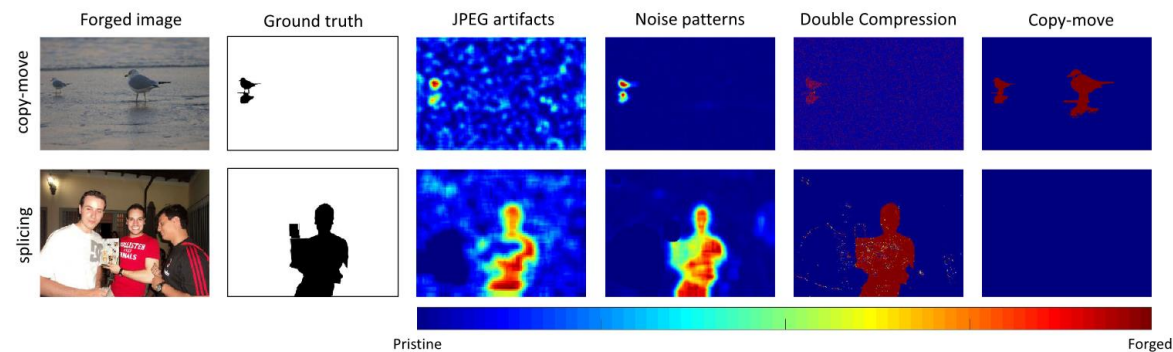
Diallo, B., Urruty, T., Bourdon, P., & Fernandez-Maloigne, C. (2020). Robust forgery detection for compressed images using CNN supervision. *Forensic Science International: Reports*, 2, 100112.

Forgery Detection Algorithms



→ Different types of forgery are detectable through different forensics traces.

→ Each forgery detection method detects a different forensics trace.

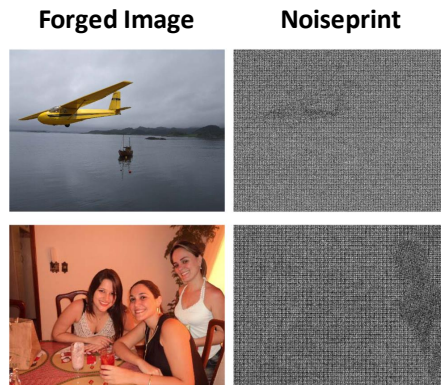


Verdoliva, L. (2020). Media forensics and deepfakes: an overview. *IEEE Journal of Selected Topics in Signal Processing*

Learnable Forensics Traces

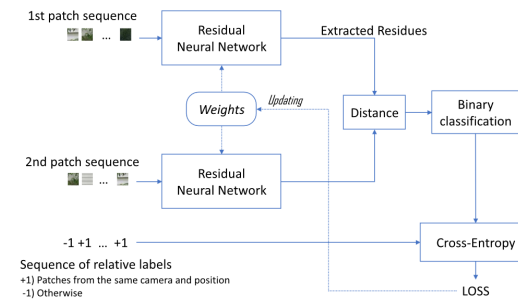


Capture multiple types of alterations at once,
without being affected by the depicted content.



Contrastive training of the network

- Generate different patterns for images from different cameras.
- Generate the same pattern for images from the same camera.



Cozzolino, D., & Verdoliva, L. (2019). Noiseprint: A CNN-based camera model fingerprint. *IEEE Transactions on Information Forensics and Security*

Learnable Forensics: The data problem

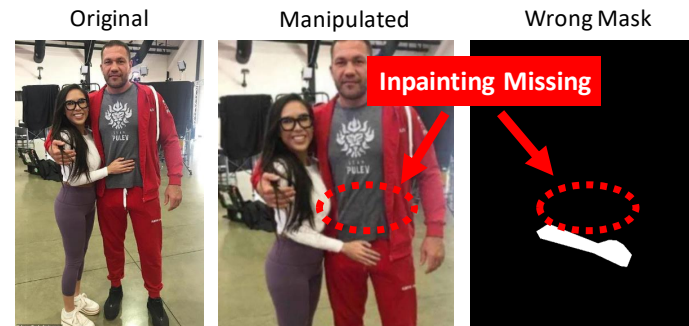


Most of the popular publicly available datasets are either outdated, wrong, or insufficient.

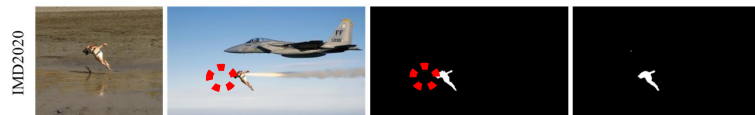
Image editing software constantly evolves (e.g. recent Adobe Photoshop includes a JPEG Artifact Removal filter, Generative Fill etc.)

Purpose-built synthetic data are commonly employed for training, covering a small portion of real-world complexity.

Most research works are trained and evaluated on suboptimal data, thus suffering when being deployed in the wild!



Harmful real-world forged image from the dataset IMD2020



Wrong-mask sample utilized as a good-example in a recent work

Hao, J., Zhang, Z., Yang, S., Xie, D., & Pu, S. (2021). Transforensics: image forgery localization with dense self-attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 15055-15064)

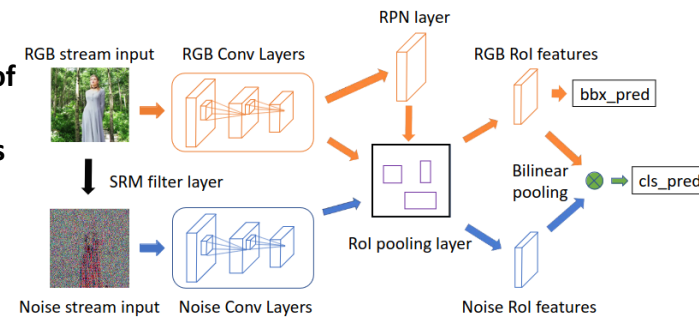
Feature-level Fusion: The current trend



- Independent network streams capture different types of forensics traces.
- Non-RGB inputs are generated through some kind of signal processing.
- Today, most state-of-the-art methods belong to this category.

Common streams in recent works

- **RGB Streams:** Capture visible artifacts (contrast differences, edge artifacts).
- **Noise Streams:** Capture noise inconsistencies (SRM filtering, High-pass filtering, Bayar Convolution).
- **DCT Streams:** Capture compression traces (DCT transform).

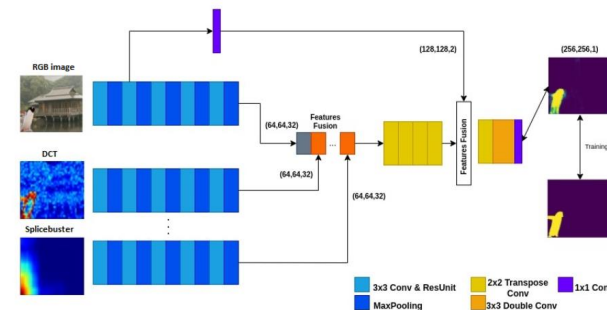


Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2018). Learning rich features for image manipulation detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*

Deploying in the wild: Algorithm-level Fusion



- Despite progress in feature fusion approaches, they are still designed with specific manipulation traces in mind e.g. only JPEG compression traces or a specific type of noise traces.
- Some methods are designed for robustness under specific scenarios. e.g. robustness against online sharing.
- Detecting image forgery in the wild requires combining the strong attributes of each method.



Algorithm-level fusion methods combine the benefits of different detection approaches into a single robust method!

Siopi, M., Kordopatis-Zilos, G., Charitidis, P., Kompatsiaris, I., & Papadopoulos, S. (2023, March). A Multi-Stream Fusion Network for Image Splicing Localization. In *MultiMedia Modeling: 29th International Conference, MMM 2023*

Image Provenance Identification



Knowing the origin of an image is a crucial step towards verifying its authenticity

- Reveals part of its processing history.
- Allows selecting suitable image forensics tools.

Multiple methods have been proposed in the past years

- Under lab conditions, some exhibit accuracies close to 100%. (Moreira et al, 2022)



Limitation: Most target the closed-set scenario!

- In the wild & at scale means a great amount of completely unknown sources!

Image Provenance Identification

In the wild & at scale



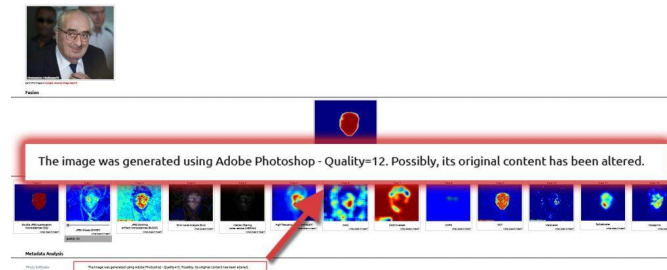
Big companies patent a lot: Usage of patented approaches is limited to specific products

JPEG Quantization Tables: Social media, camera makers, software companies have their own approaches

JPEG Quantization Tables Database: Over 4000 quantization tables of known origin

Patented Compression Patterns → Low false-detection rate

Seamless integration with most published provenance identification methods that use JPEG QTs



<https://mever.gr/2022/11/09/identifying-image-provenance-in-the-wild/>

Image forensics at scale?




Most image forensics algorithms and their available implementations are:

→ **Extremely slow** when applied on big images.

→ **Non-maintained** and built upon out-of-date libraries.

→ **Inadequate for use at-scale** and intended for one-time research-oriented evaluations.

Greatly optimized the efficiency of multiple popular forgery detection algorithms	
Algorithmic Improvements	Implementation Improvements
New Distributed and Scalable Architecture	
	

Algorithm	Speedup
BLK	x4.5
CAGI	x9
Mantranet	x35
Median	x2
Wavelet	x200
Splicebuster	x3.5
Noiseprint	x3
SPAN	x8

Summary & Challenges



Results of Image Forensics Research & Software Engineering MAAM Image Verification Assistant

- **Quick & Robust High-Level Reasoning:** Fusion
- **Interpretable Results:** 15-level forensics insights & provenance identification
- **Optimized for large-scale verification!**

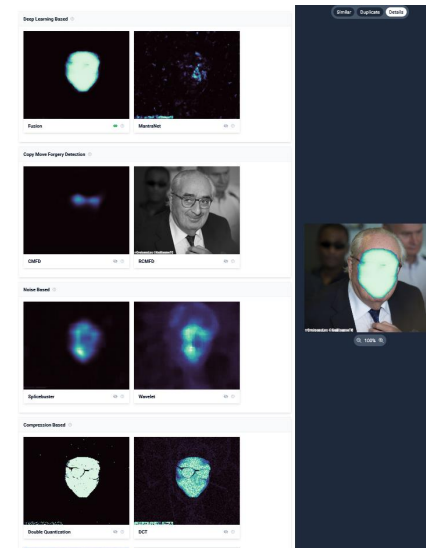
But as image editing software evolves:

Learning based algorithms & non-public data!



Fragmentation & research not applicable in real-world conditions!

Up-to-date common evaluation standards are required!



Dimitrios Karageorgiou

CERTH



Consortium Partners



CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS



Atos



UAB
Universitat Autònoma
de Barcelona

SWISS TXT



TIMELEX



<https://mediaverse-project.eu/>

MediaVerse: A universe of media assets and
co-creation opportunities at your fingertips

Co-financed by the EC
under Grant agreement
ID: 957252

