

La preuve en image ? Mon œil !

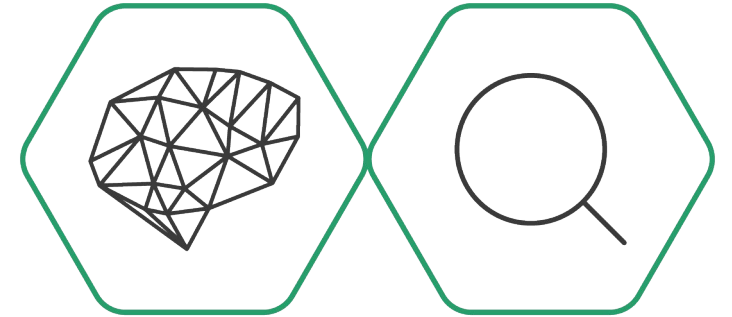
Apprenez à détecter les images falsifiées

Quentin Bammey

ENS Paris-Saclay, Centre Borelli

30 mars 2023

Atelier "Décrypter l'info"



vera.ai



CENTRE
BORELLI

école
normale
supérieure
paris-saclay

université
PARIS-SACLAY

Les images falsifiées sont partout : humour



<https://twitter.com/GuillaumeTC>

#CroisonsLes @GuillaumeTC

Les images falsifiées sont partout : fake news

18 déc. 2019

Elle n' est pas belle la France ?
Photo prise à la **CAF de Rosny sous Bois**.
D'après la loi en vigueur sur le **voile**, elles sont toutes en infraction, mais elles auront quand même leur chèque. Continuez à travailler dur car elles ont besoin de votre argent et de votre soutien.



🗨️ ↻️ ❤️ 📊 ↗️

Les images falsifiées sont partout : fake news



La photo a été prise à Londres, et a été falsifiée pour la détourner de son contexte !

Comment savoir si une image est falsifiée ?

Recherche inversée



- Google Images, TinEye

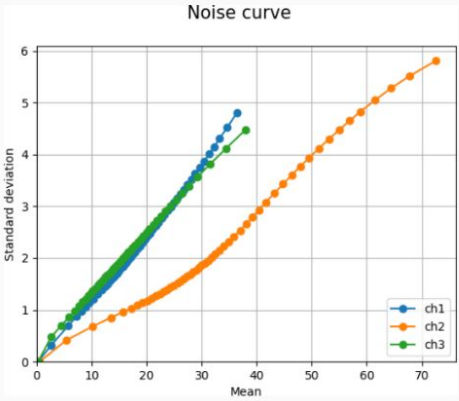
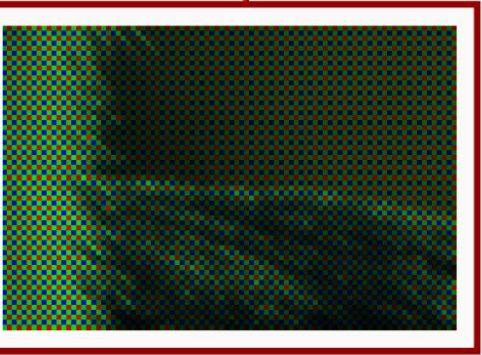
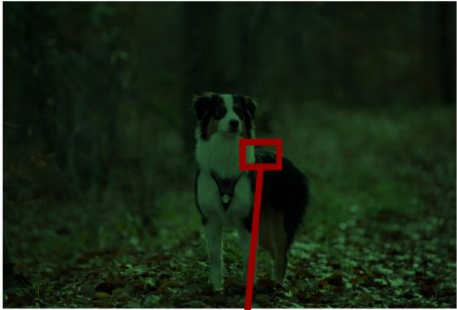


A screenshot of the TinEye search results page. The page header includes the TinEye logo, navigation links for Search, Technology, Products, and About, and a "We are hiring" button. The search bar contains the URL "https://pbs.twimg.com/media/DESrS1gXcAU-_MQ?format=jpg&name=" and a magnifying glass icon. Below the search bar, there is a section for "636 results" with a sub-header "Searched over 59.2 billion images in 0.7 seconds for: pbs.twimg.com/media/DESrS1gXcAU-_MQ?format=jpg&name=s...". There is a checkbox for "Include 7 results not available" which is checked. Below the results, there are filters for "Sort by biggest image" and "Filter by website / collection". The top result is from "www.reddit.com" with the title "/pics/comments/6nzdj/when_youre_t..." and the text "- First found on Sep 21, 2017". The filename is "cumhurbaskani-erdogan-g20-zirvesinde-boyle-karsilandi-1499..." and the size is "(2047 x 1262, 747.6 kB)".

La vie secrète d'une image



Raw
acquisition



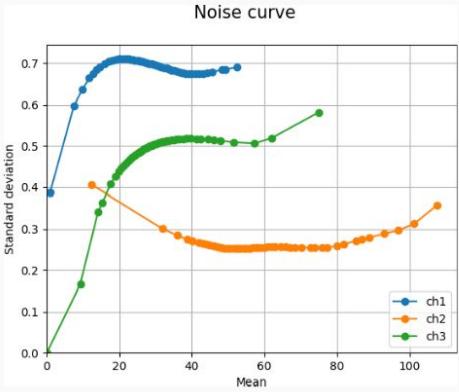
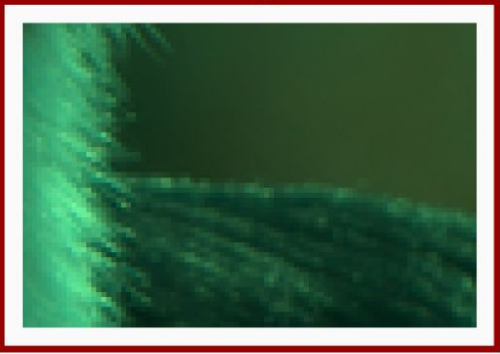
La vie secrète d'une image



Raw acquisition



Demosaicing



La vie secrète d'une image



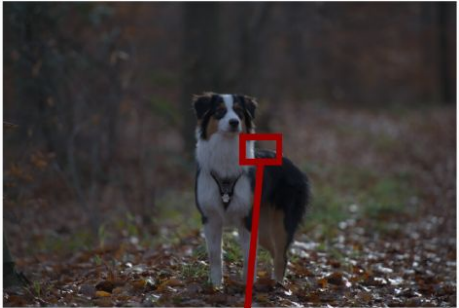
Raw acquisition



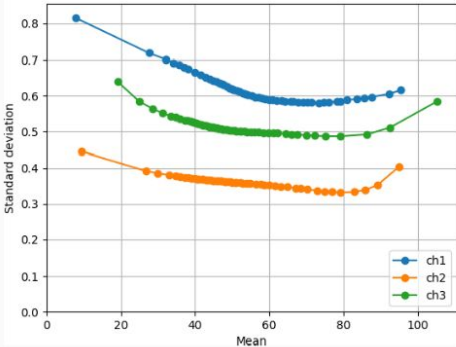
Demosaicing



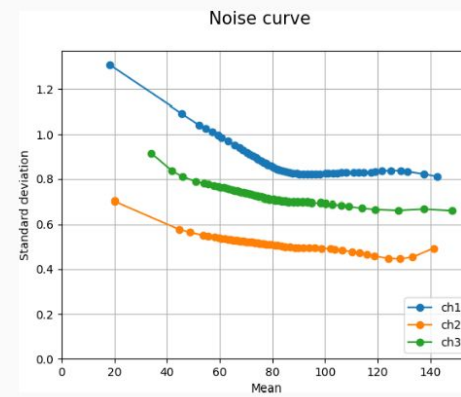
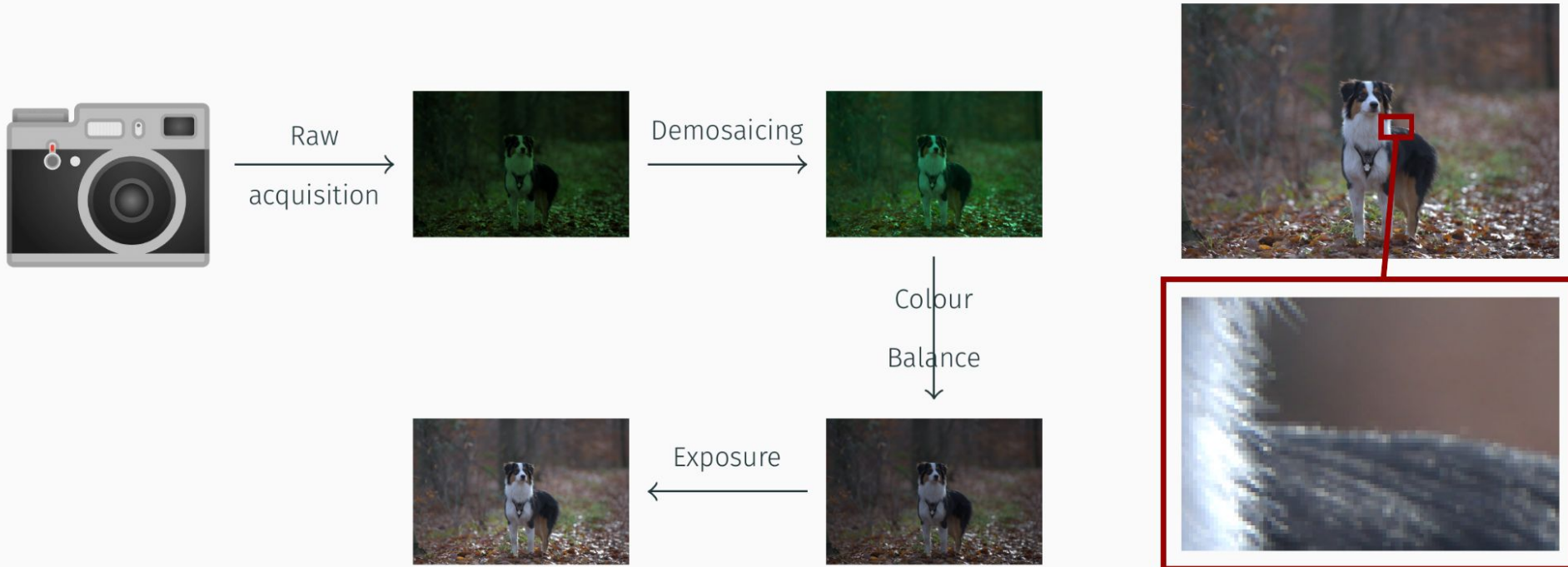
Colour Balance



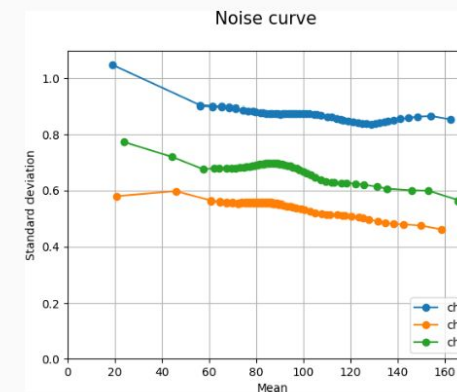
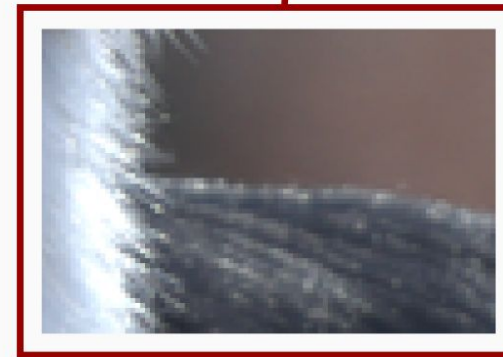
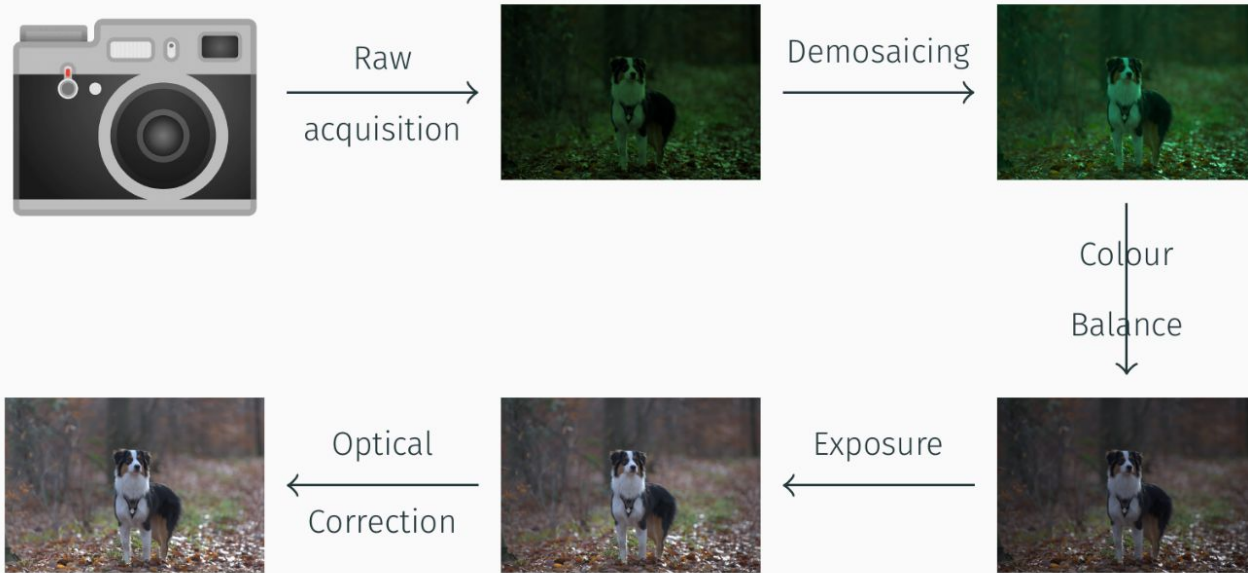
Noise curve



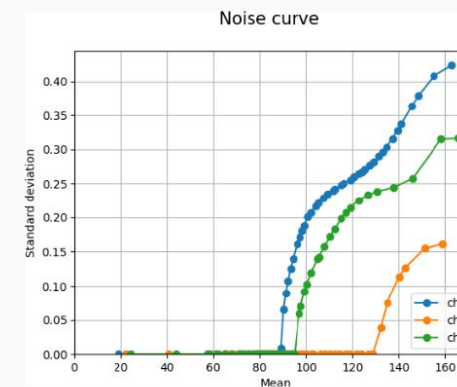
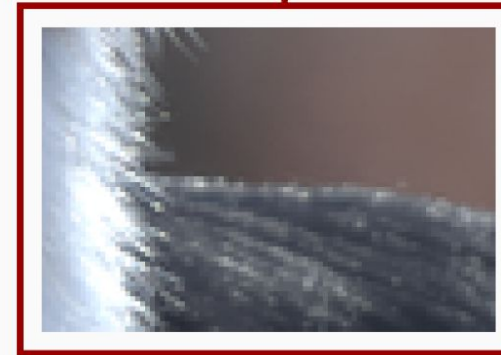
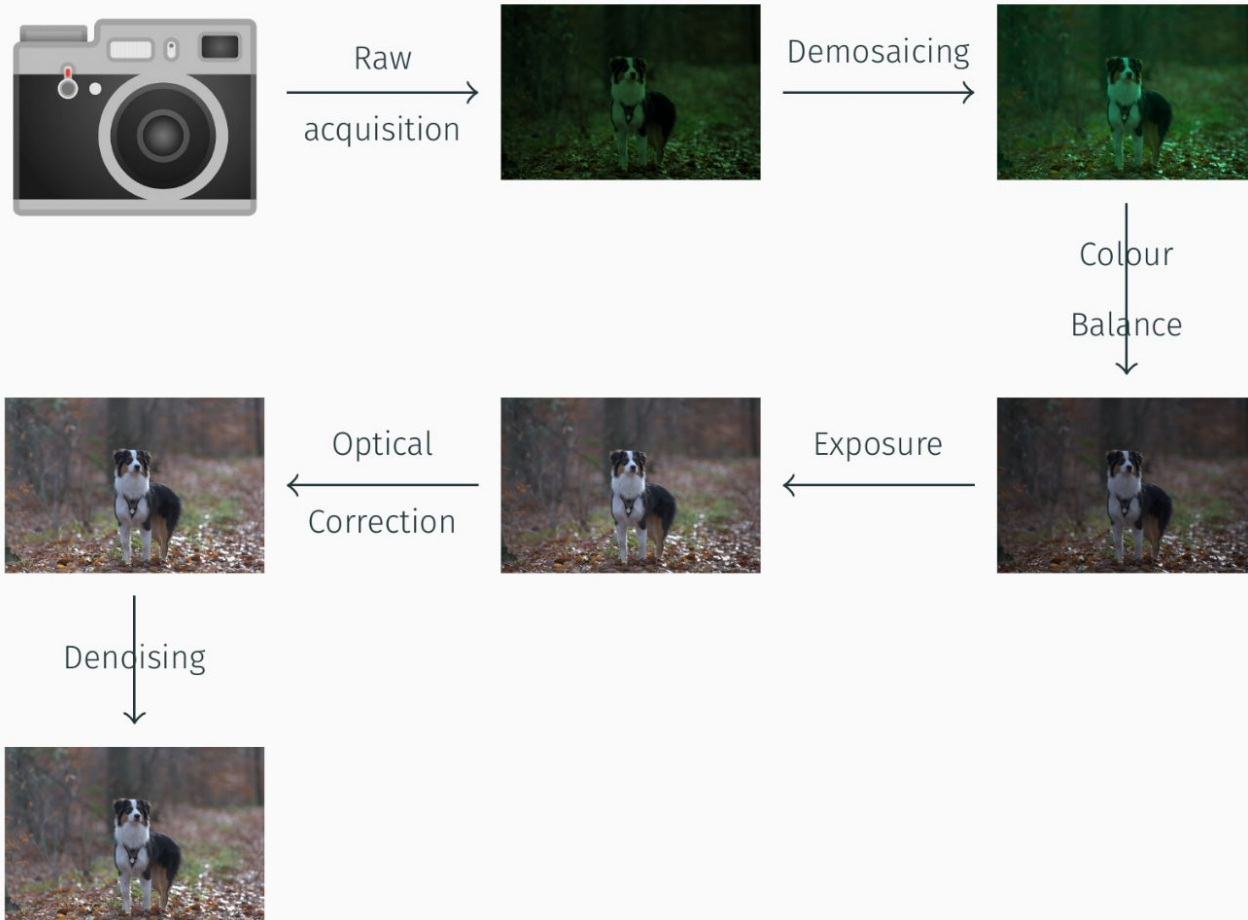
La vie secrète d'une image



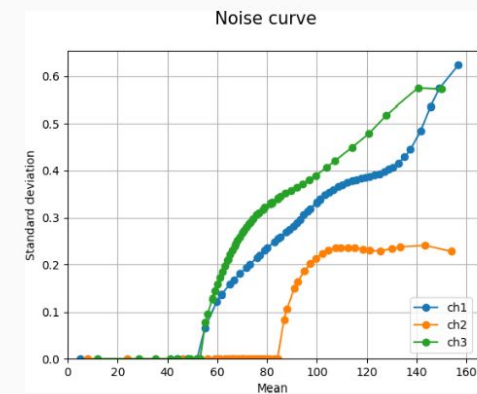
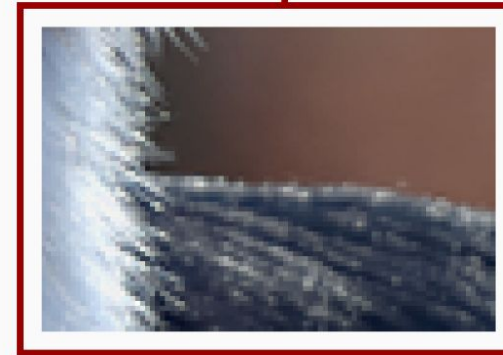
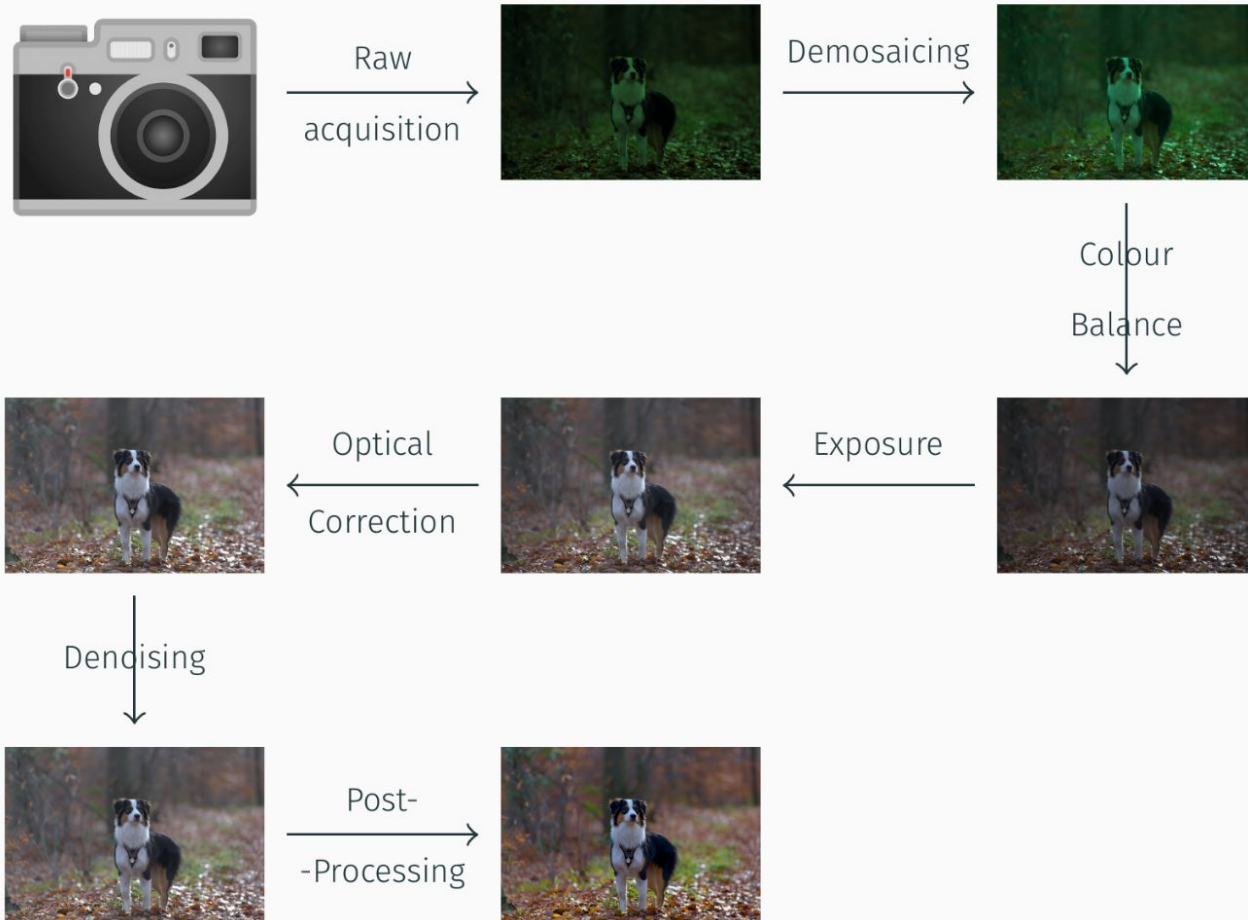
La vie secrète d'une image



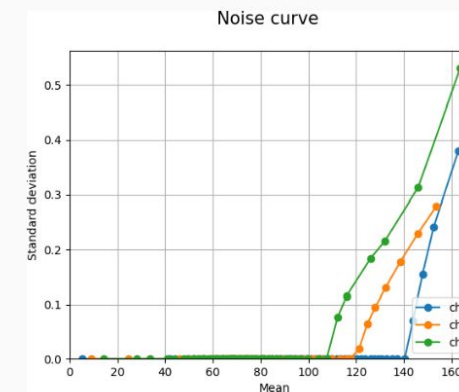
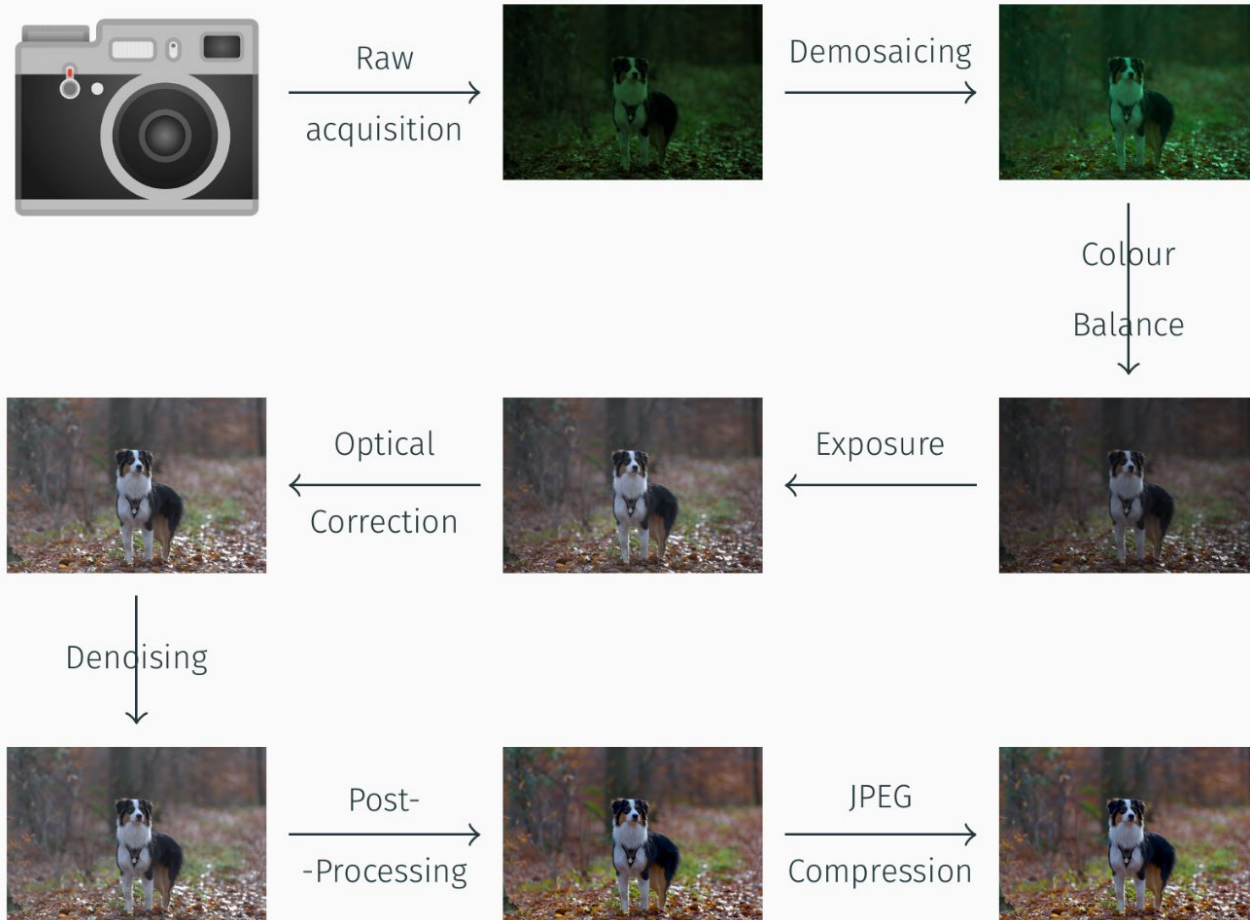
La vie secrète d'une image



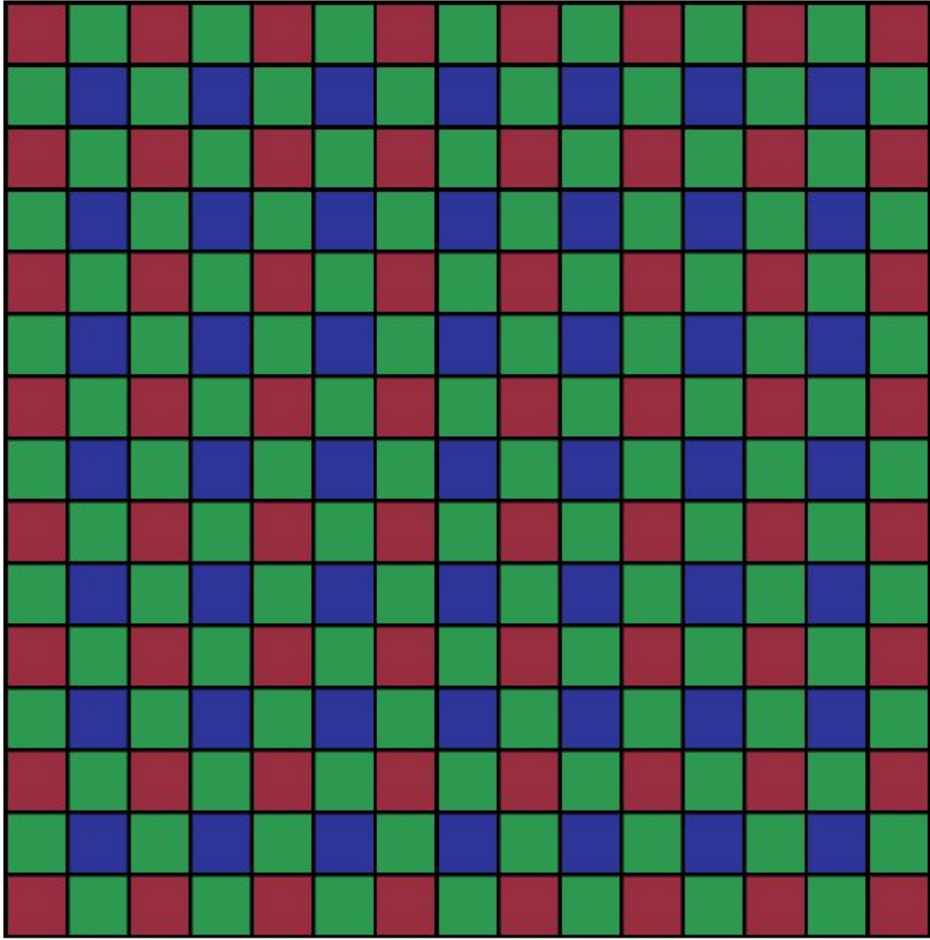
La vie secrète d'une image



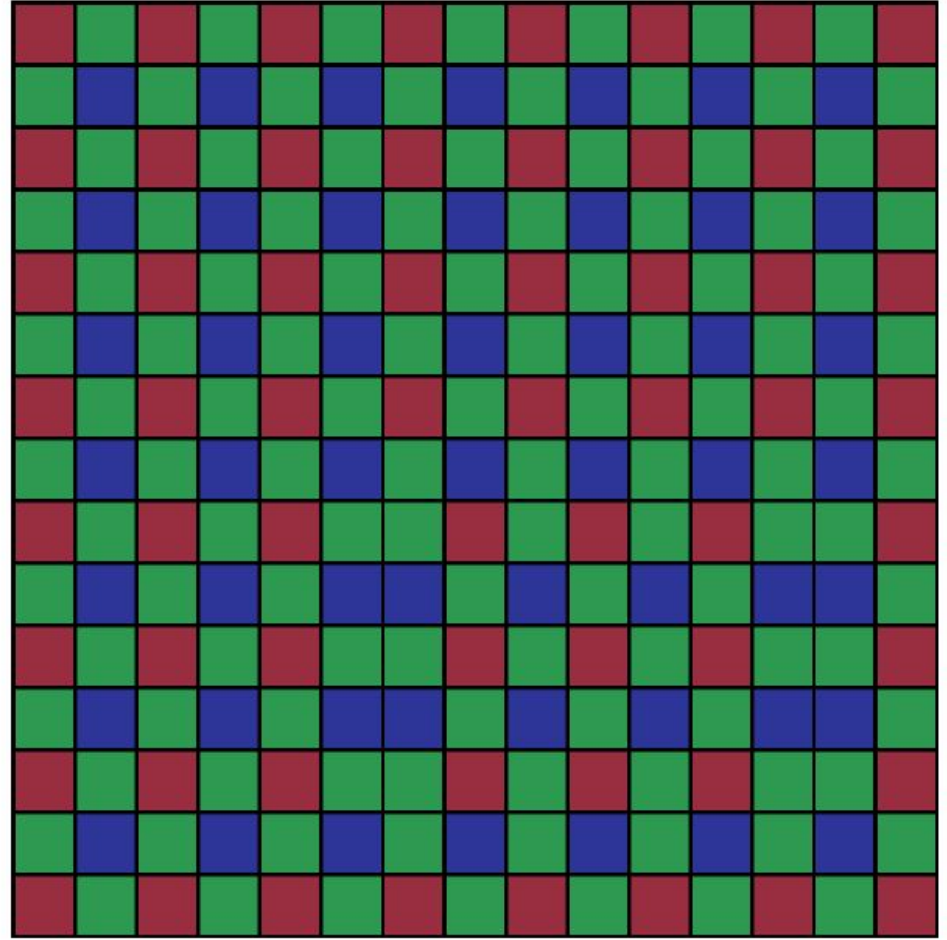
La vie secrète d'une image



Détection de falsifications à travers l'analyse de démosaïquage



(a) Authentic image



(b) Forged image

Détection de falsifications à travers l'analyse de démosaïquage



(a) Forged image



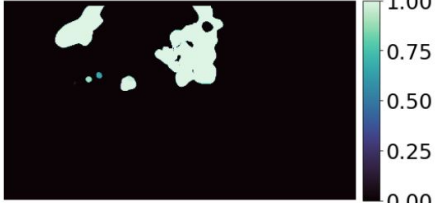
(b) Original image



(c) Forgery mask



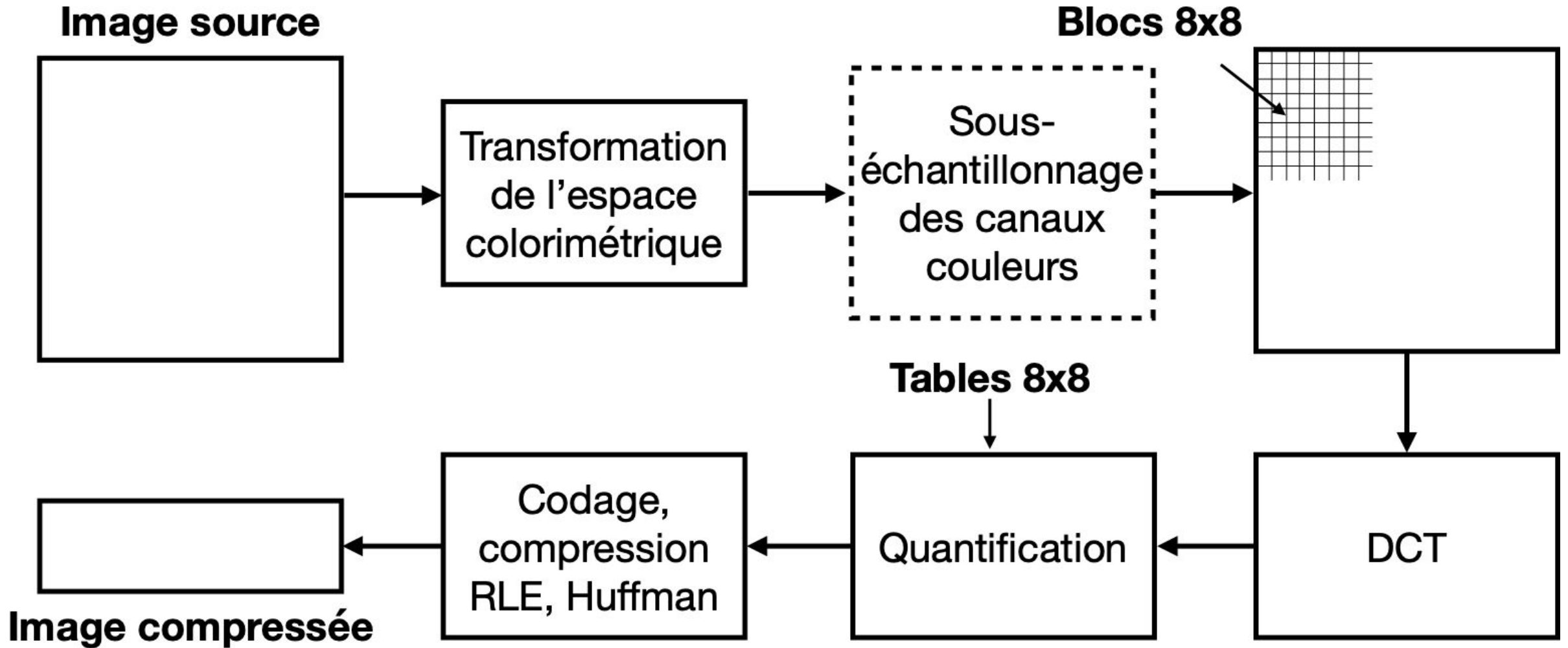
(d) Detected demosaicing grids



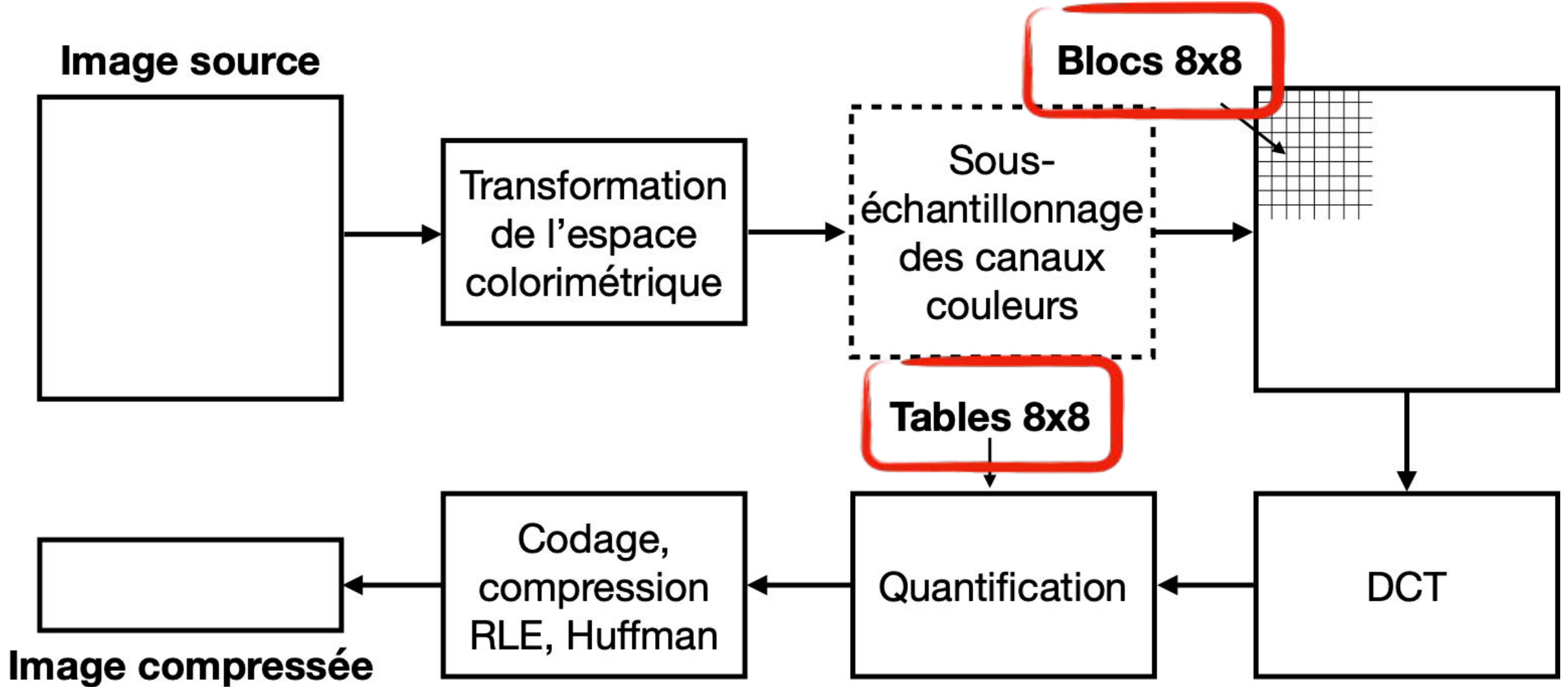
(e) Detected forgeries

Source : base de données Korus

Détection de falsifications à travers l'analyse de compression JPEG



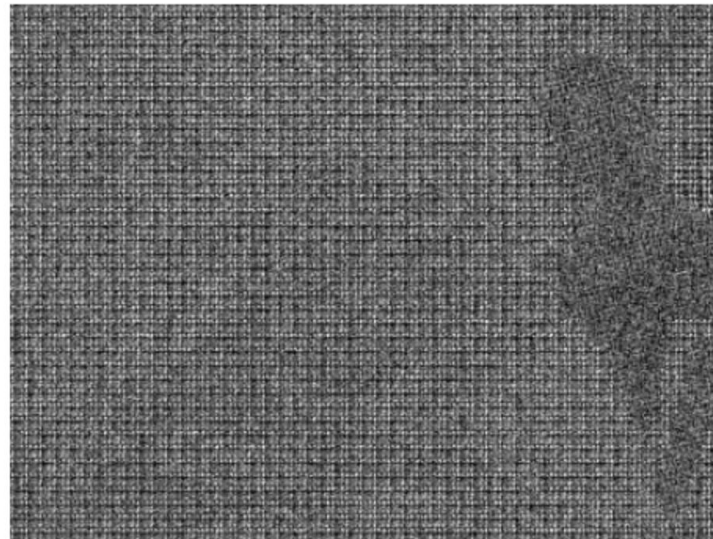
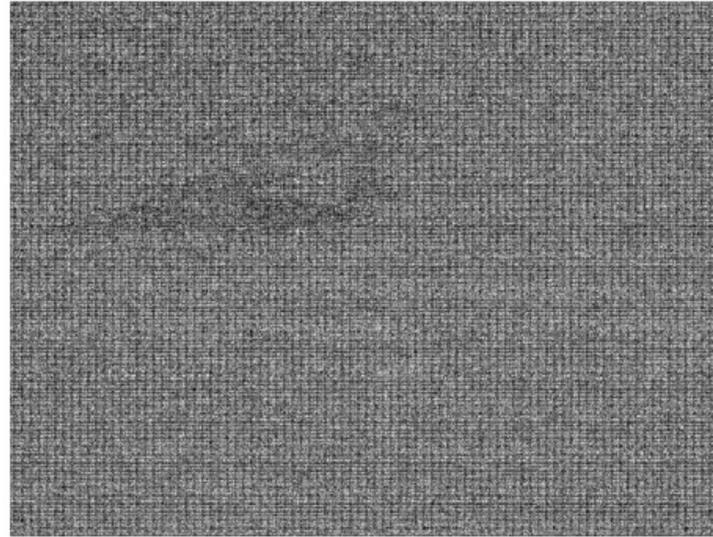
Détection de falsifications à travers l'analyse de compression JPEG



Détection de falsifications à travers l'analyse de compression JPEG



Détection de falsifications grâce aux textures



Cozzolino, D., & Verdoliva, L. (2019). Noiseprint: A CNN-based camera model fingerprint. *IEEE Transactions on Information Forensics and Security*, 15, 144-159.

Détection de falsifications à travers l'analyse de bruit



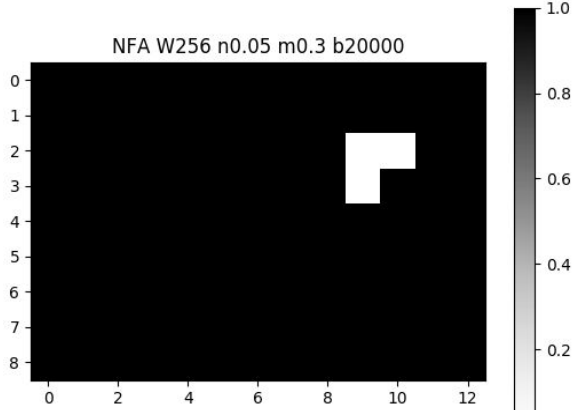
Original image



Face Swap



Detection



Statistical validation

Détection de copier-coller

Objets similaires mais différents

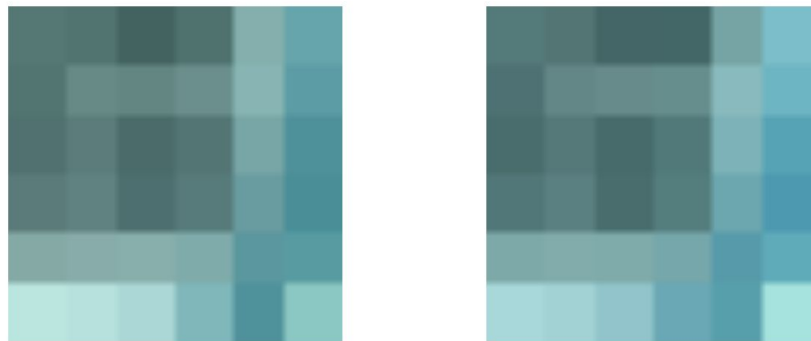


Figure 1: Deux descripteurs similaires (dont les positions sont représentés par un point rouge) de deux objets similaires: Les descripteurs ne sont pas identiques.

Détection de copier-coller

Copies falsifiées



Figure 2: Deux descripteurs similaires (dont les positions sont représentés par un point rouge) d'une région clonée: Les descripteurs sont identiques.

Détection de copier-coller



Figure 3: Objets similaires mais différents (base COVERAGE)



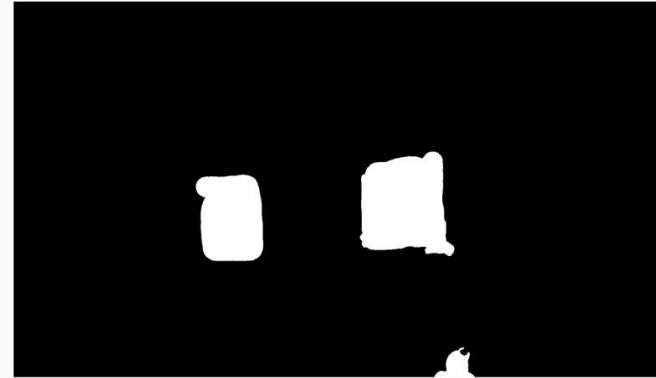
Figure 4: Image falsifiée (base COVERAGE)

Comment détecter une falsification ?

Des méthodes obscures...



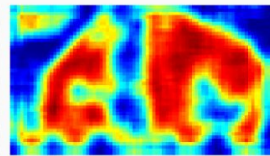
Image



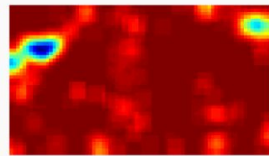
Ground Truth



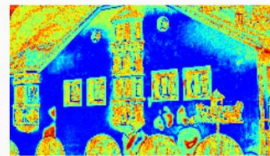
Noiseprint



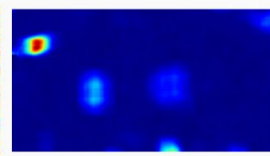
CAGI



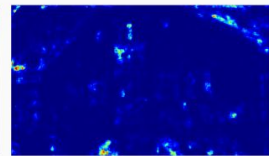
CAGI-Inv



DQ



Splicebuster



ManTraNet

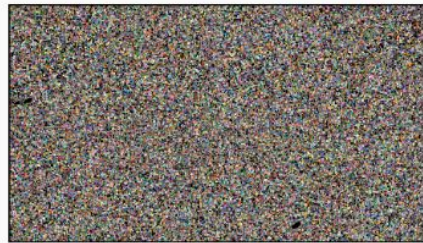
Beaucoup de méthodes existantes nécessitent d'être interprétées :

- Quelles parties correspondent vraiment à une falsification ?
- Quelles méthodes sont fiables ?
- Sur quelles justifications basent-elles leur détection ?

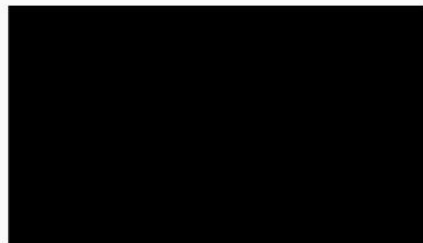
Comment détecter une falsification ? ... Vers des méthodes interprétables



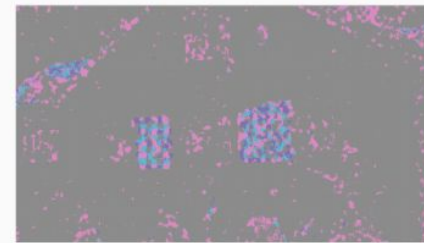
Ground Truth



ZERO ↓



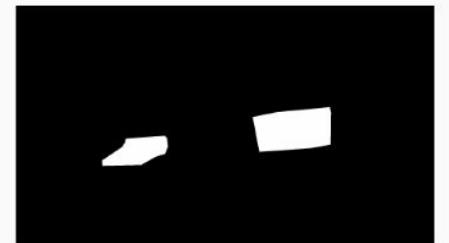
Noise ↓ sniffer



CFA ↓

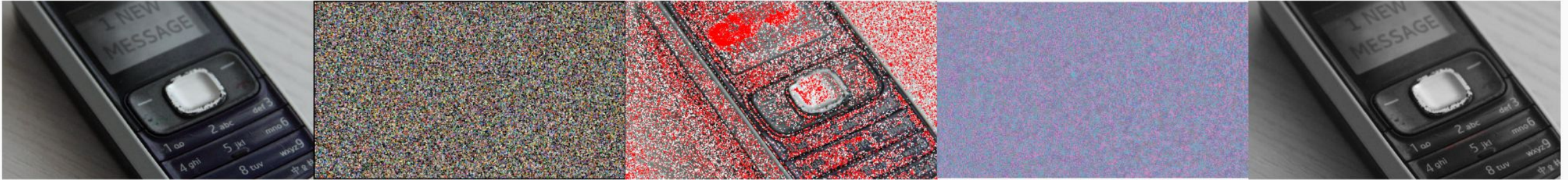


RCMFD ↓



Objectif : Créer des méthodes robustes, qui analysent elles-mêmes leur sorties pour obtenir une détection automatique et justifiée

Comment détecter une falsification ? ... Vers des méthodes interprétables



Image

Zero

Noisesniffer

CFA

RCMFD



Pas de détection avec une méthode = Un type d'incohérence n'est pas présent.

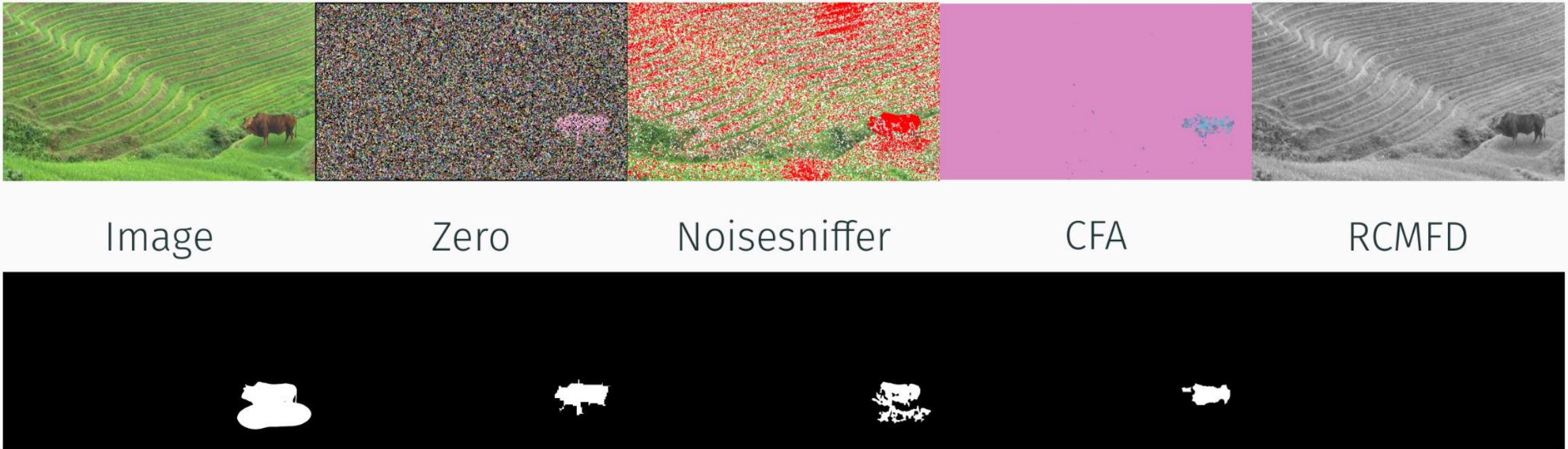
Ça ne veut pas dire que l'image est authentique !

Ici :

- ZERO ne détecte rien car l'image n'est pas compressée JPEG : la méthode détecte simplement que l'image n'a pas de traces de compression
- CFA ne détecte rien car l'image est en noir est blanc
- RCMFD ne détecte rien = pas de copier-coller visible

Mais on voit dans l'analyse des niveaux de bruit que l'image est bien falsifiée !

Comment détecter une falsification ? ... Vers des méthodes interprétables



Pas de détection avec une méthode = Un type d'incohérence n'est pas présent.
Ça ne veut pas dire que l'image est authentique !

Ici :

- RCMFD ne détecte rien = pas de copier-coller visible

Mais on voit dans l'analyse des niveaux de bruit que l'image est bien falsifiée !


Comment utiliser des outils de détection de falsification ?



- InVID-WeVerify : Un plugin de vérification d'images et bien plus
 - <https://www.invid-project.eu/tools-and-services/invid-verification-plugin/>
 - Co-développé par l'Agence France Presse, le Centre Borelli (ENS Paris-Saclay), et l'ITI CERTH
 - Contient aussi des ressources d'enseignement
- Démonos IPOL
 - IPOL : Journal scientifique en ligne
 - Chaque article est accompagné d'une démo utilisable par tous
 - Exemples :
 - ZERO : <http://www.ipol.im/pub/art/2021/390/>
 - CFA (Ancienne version) : <http://www.ipol.im/pub/art/2021/355/>
 - Détection de copier-coller : <http://www.ipol.im/pub/art/2018/213/>
 - ManTraNet : <http://www.ipol.im/pub/art/2022/431/>




Comment utiliser des outils de détection de falsification ?

InVIDTOOLSASSISTANTTUTORIALDEMOCLASSROOMABOUTEnglish🌐

⚠️ This enhanced forensic toolkit aims to help you detect alterations in manipulated images. You should avoid using it with screenshots, scanned images of documents, or juxtaposed images that are in fact altered images. More filters are highlighting the same zone, more suspicious is that particular area of the image. Please take into account that forensic filters are outlining any digital signal alteration and not only semantically manipulated artefacts (which means false positives are possible). Some complex textures or excess of luminance may also alter the signal without any manipulation intention. Whenever possible, use systematically the best image resolution available (even by searching through similarity for higher resolution identical images).

Analysed Image



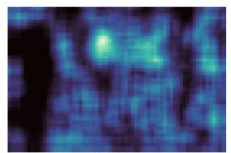

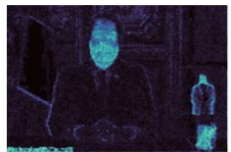
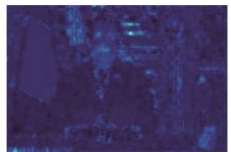


#CroisLes @GuillaumeTC

NEW IMAGE

Filters

COMPRESSION TRACES DEEP LEARNING CLONING

 <p>Zero</p>	 <p>GHOST</p>	 <p>CAGI</p>
 <p>Double Quantization</p>	 <p>DCT</p>	 <p>BLOCK</p>




No detection Detection

🔍 Mouse over the filters to see a transparent mask with the results on the image.

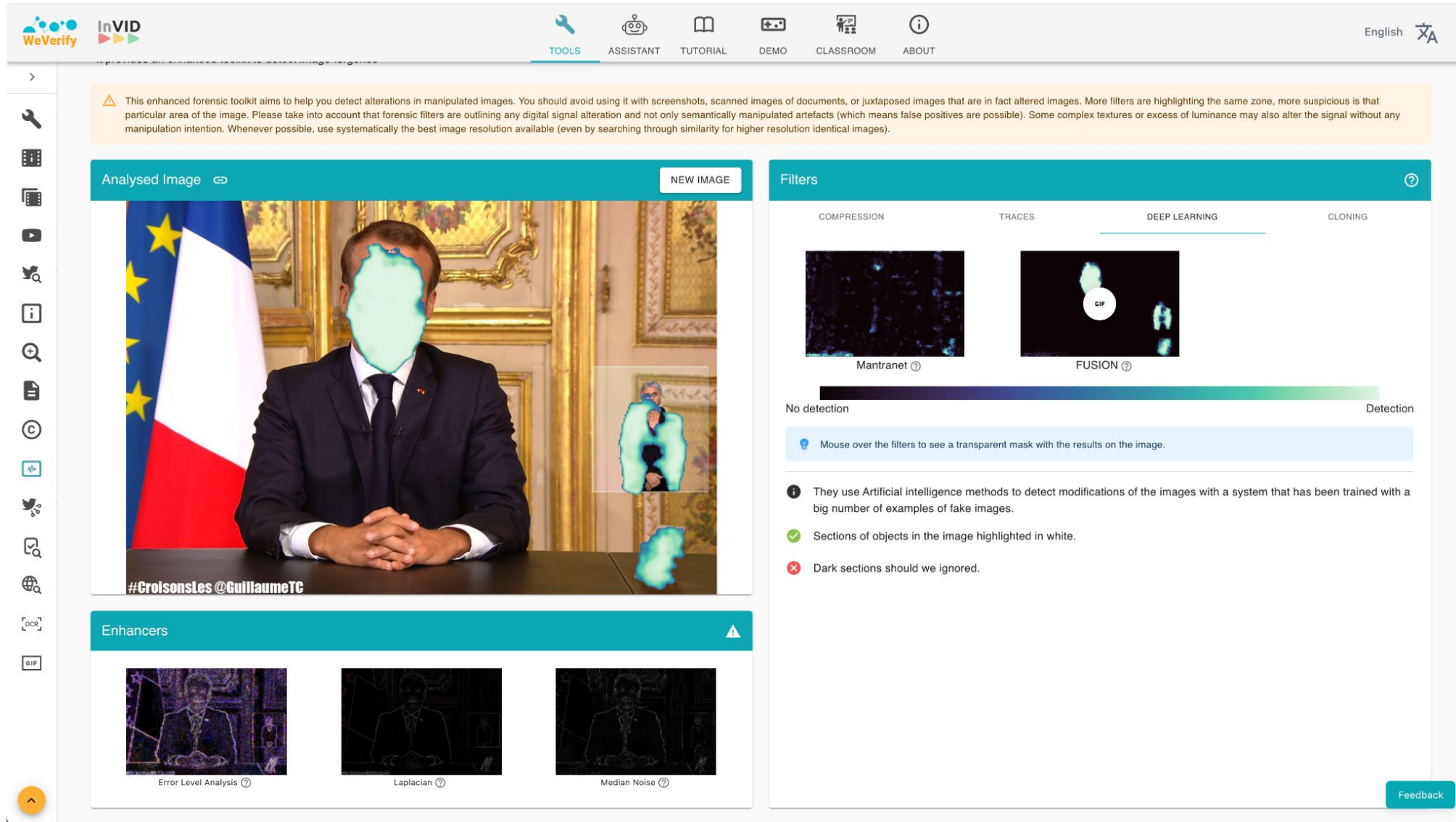
- 📌 Those filters detect anomalies in previous processes of creation and compression of the image. If a new element has been removed or added to the image, it can be detected if it has a different compression than the rest of the image.
- ✅ Combination of multiple filters that outline in white or green-light a common area of the image.
- ❌ Complex textures of an object or saturated areas of the photo can generate false positives.

Feedback

Enhancers

 <p>Error Level Analysis</p>	 <p>Laplacian</p>	 <p>Median Noise</p>
---	--	--

Comment utiliser des outils de détection de falsification ?



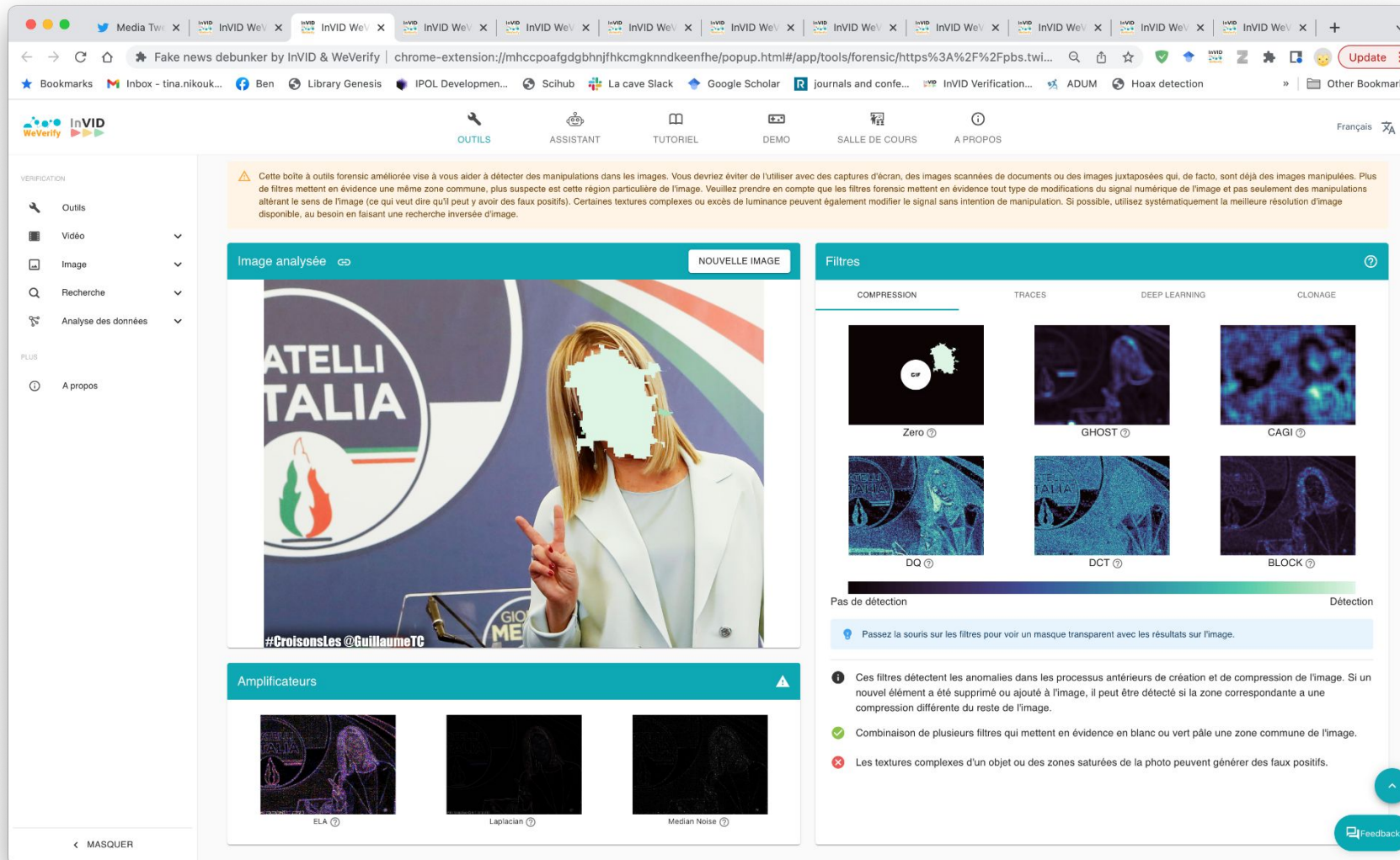
The screenshot shows the WeVerify InVID web interface. At the top, there is a navigation bar with icons for TOOLS, ASSISTANT, TUTORIAL, DEMO, CLASSROOM, and ABOUT, along with a language selector set to English. A warning message is displayed at the top of the main area, advising users to avoid using the tool on screenshots or scanned documents and to use the highest resolution available.

The main interface is divided into several sections:

- Analysed Image:** Displays a video frame of Emmanuel Macron with a cyan-colored detection mask over his face. A smaller inset shows a zoomed-in view of the mask. The text "#CroisonsLes @GuillaumeTC" is visible at the bottom of the image.
- Filters:** A panel on the right showing detection results for different filters. The "DEEP LEARNING" filter is active, showing a "FUSION" result with a white circle labeled "CIF" on the face. Other filters shown include "Mantranet". A color scale at the bottom of this panel ranges from "No detection" (black) to "Detection" (white).
- Enhancers:** A panel at the bottom showing three different image enhancement filters: "Error Level Analysis", "Laplacian", and "Median Noise".

Additional features include a sidebar on the left with various tool icons and a "Feedback" button in the bottom right corner.

Comment utiliser des outils de détection de falsification ?



The screenshot displays the InVID WeVerify web application interface. At the top, there is a navigation bar with the InVID logo and various menu items: Outils, ASSISTANT, TUTORIEL, DEMO, SALLE DE COURS, and A PROPOS. A language selector for 'Français' is also present.

The main content area is divided into several sections:

- VERIFICATION:** A sidebar on the left lists navigation options: Outils, Vidéo, image, Recherche, Analyse des données, and A propos.
- Image analysée:** The central area shows an image of a woman with a redacted face. The image is overlaid with a green grid, indicating the analysis process. A 'NOUVELLE IMAGE' button is located at the top right of this section.
- Filtres:** A panel on the right offers various detection filters: COMPRESSION (Zero, DQ, DCT), TRACES (GHOST), DEEP LEARNING (CAGI), and CLONAGE (BLOCK). A color scale at the bottom of this panel ranges from 'Pas de détection' (black) to 'Détection' (red).
- Amplificateurs:** A section below the main image shows three zoomed-in areas: ELA, Laplacian, and Median Noise.

At the bottom right, there is a 'Feedback' button and a small upward arrow icon.

Comment utiliser des outils de détection de falsification ?



Accueil

Explorer

Notifications

Messages

Signets

Listes

Profil

Plus

Tweeter



GuillaumeTC
11,9 k photos et vidéos

Abonné



GuillaumeTC @GuillaumeTC · 12 déc. 2022
Quand je pense à Mercredi.

#CroisonsLes



#CroisonsLes @GuillaumeTC

39 138 1107



GuillaumeTC @GuillaumeTC · 10 déc. 2022
ET POURQUOI ON VEUT LA COUPE DU MONDE ?
PARCE QUE C'EST NOTRE PROJEEEEEEEEET !

#CroisonsLes #AngFra #EngFra

Recherche Twitter

Divertissement · Tendances

Paul Mirabel

Tendances

#HogwartsLegacy

8 832 Tweets

Tendances

#Russie

3 837 Tweets

Tendance dans la catégorie Ile-de-France

#RATP

1 114 Tweets

Tendance dans la catégorie France

Dutroux

1 049 Tweets

Politique · Tendances

#greve19janvier

16,2 k Tweets

Tendance dans la catégorie Ile-de-France

Bible

77,8 k Tweets

Politique · Tendances

Afrique

14,7 k Tweets

Tendance dans la catégorie France

#neige

2 685 Tweets

Tendance dans la catégorie France

Grève

2 764 Tweets

Messages



Tina
@TinaNkh

Comment utiliser des outils de détection de falsification ?



The screenshot shows a Twitter interface with a tweet from GuillaumeTC (@GuillaumeTC) dated December 12, 2022. The tweet text is "Quand je pense à Mercredi." and includes the hashtag #CroisonsLes. The image in the tweet is a portrait of Wednesday Addams. A context menu is open over the image, listing various actions. A sub-menu is also open over the "Image Forensic" option, listing several image analysis tools.

Context Menu Options:

- Open Link in New Tab
- Open Link in New Window
- Open Link in Incognito Window
- Create QR code for this Image
- Save Link As...
- Copy Link Address
- Open Image in New Tab
- Save Image As...
- Copy Image
- Copy Image Address
- Search image with Google
- AdGuard AdBlocker
- Fake news debunker by InVID WeVerify
- Windscribe - Free Proxy and Ad Blocker

Image Forensic Sub-menu Options:

- Open with assistant
- Open with OCR
- Video Reverse Search - DBKF (beta)
- Video contextual Analysis
- Image Magnifier
- Image Forensic**
- Image Reverse Search - ALL
- Image Reverse Search - DBKF (beta)
- Image Reverse Search - Google
- Image Reverse Search - Google Lens
- Image Reverse Search - Yandex
- Image Reverse Search - Bing
- Image Reverse Search - TinEye
- Image Reverse Search - Baidu
- Image Reverse Search - Reddit

Twitter Interface Elements:

- Profile: GuillaumeTC (11,9 k photos et vidéos)
- Follow button: Abonné
- Search: Recherche Twitter
- Navigation: Accueil, Explorer, Notifications, Messages, Signets, Listes, Profil, Plus
- Buttons: Tweeter
- Engagement: 39 replies, 138 retweets, 1107 likes
- Bottom tweet: GuillaumeTC (10 déc. 2022) "ET POURQUOI ON VEUT LA COUPE DU MONDE ? PARCE QUE C'EST NOTRE PROJEEEEEEEEET !"

Comment utiliser des outils de détection de falsification ?

Forensic

Outils Avancés

DECONNEXION

Les outils avancés sont déverrouillés

Une boîte à outils améliorée pour détecter les falsifications d'images

⚠ Cette boîte à outils forensic améliorée vise à vous aider à détecter des manipulations dans les images. Vous devriez éviter de l'utiliser avec des captures d'écran, des images scannées de documents ou des images juxtaposées qui, de facto, sont déjà des images manipulées. Plus de filtres mettent en évidence une même zone commune, plus suspecte est cette région particulière de l'image. Veuillez prendre en compte que les filtres forensic mettent en évidence tout type de modifications du signal numérique de l'image et pas seulement des manipulations altérant le sens de l'image (ce qui veut dire qu'il peut y avoir des faux positifs). Certaines textures complexes ou excès de luminosité peuvent également modifier le signal sans intention de manipulation. Si possible, utilisez systématiquement la meilleure résolution d'image disponible, au besoin en faisant une recherche inversée d'image.

Image analysée

NOUVELLE IMAGE

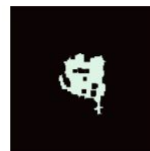
Filtres

COMPRESSION

TRACES

DEEP LEARNING

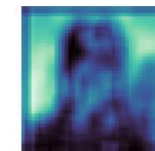
CLONAGE



Zero



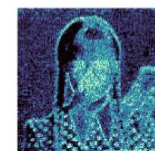
GHOST



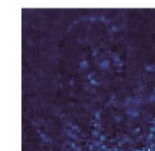
CAGI



DQ



DCT



BLOCK

Pas de détection

Détection

👉 Passez la souris sur les filtres pour voir un masque transparent avec les résultats sur l'image.

ⓘ Ces filtres détectent les anomalies dans les processus antérieurs de création et de compression de l'image. Si un nouvel élément a été supprimé ou ajouté à l'image, il peut être détecté si la zone correspondante a une compression différente du reste de l'image.

✅ Combinaison de plusieurs filtres qui mettent en évidence en blanc ou vert pâle une zone commune de l'image.

❌ Les textures complexes d'un objet ou des zones saturées de la photo peuvent générer des faux positifs.

Amplificateurs



Feedback

Comment utiliser des outils de détection de falsification ?

Forensic

Outils Avancés

DECONNEXION

Les outils avancés sont déverrouillés

Une boîte à outils améliorée pour détecter les falsifications d'images

⚠ Cette boîte à outils forensic améliorée vise à vous aider à détecter des manipulations dans les images. Vous devriez éviter de l'utiliser avec des captures d'écran, des images scannées de documents ou des images juxtaposées qui, de facto, sont déjà des images manipulées. Plus de filtres mettent en évidence une même zone commune, plus suspecte est cette région particulière de l'image. Veuillez prendre en compte que les filtres forensic mettent en évidence tout type de modifications du signal numérique de l'image et pas seulement des manipulations altérant le sens de l'image (ce qui veut dire qu'il peut y avoir des faux positifs). Certaines textures complexes ou excès de luminosité peuvent également modifier le signal sans intention de manipulation. Si possible, utilisez systématiquement la meilleure résolution d'image disponible, au besoin en faisant une recherche inversée d'image.

Image analysée ↔

NOUVELLE IMAGE



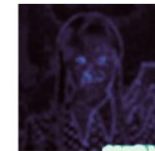
#CroisonsLes @GuillaumeTC

Filtres

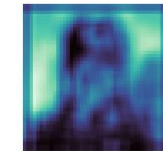
COMPRESSION TRACES DEEP LEARNING CLONAGE



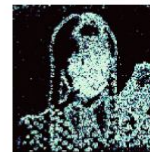
Zero



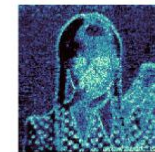
GHOST



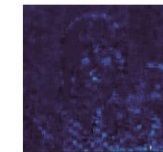
CAGI



DQ



DCT



BLOCK

Pas de détection

Détection

ℹ Passez la souris sur les filtres pour voir un masque transparent avec les résultats sur l'image.

ℹ Ces filtres détectent les anomalies dans les processus antérieurs de création et de compression de l'image. Si un nouvel élément a été supprimé ou ajouté à l'image, il peut être détecté si la zone correspondante a une compression différente du reste de l'image.

✓ Combinaison de plusieurs filtres qui mettent en évidence en blanc ou vert pâle une zone commune de l'image.

✗ Les textures complexes d'un objet ou des zones saturées de la photo peuvent générer des faux positifs.

Feedback

VERIFL...



PLUS



Quel futur pour les falsifications ? Des deepfakes...

<https://youtu.be/cQ54GDm1eL0>



Quel futur pour les falsifications ? ... aux images de synthèse



Quel futur pour les falsifications ? ... aux images de synthèse



Quel futur pour les falsifications ? ... aux images de synthèse



Quel futur pour les falsifications ? ... aux images de synthèse



Comment les détecter ?

- Deepfakes : quelques outils en ligne, peu fiables
- Images de synthèse : Trop récent, pas encore d'outils disponibles
- Inspection visuelle :
 - Des textures souvent plates (visage sans imperfection, aucun bruit dans les images...)
 - Problèmes de symétrie (visage asymétrique, yeux qui louchent ou ne sont pas de la même couleur, etc)
 - Détails erronés (Regarder les mains, chercher des incohérences dans l'arrière-plan, etc)
 - Vidéos : Les mouvements de visage ne semblent pas toujours naturels
- Attention ! Les outils de génération s'améliorent : Les images vont devenir de plus en plus difficiles à authentifier visuellement

Quentin Bammey

Contact: quentin.bammey@ens-paris-saclay.fr



Follow us on Twitter: [@veraai_eu](https://twitter.com/veraai_eu)

Website: <https://www.veraai.eu/>

Co-financed by the European Union, Horizon Europe programme, Grant Agreement No 101070093.

Additional funding from Innovate UK grant No 10039055 and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract No 22.00245

