La preuve en image ? Mon œil !

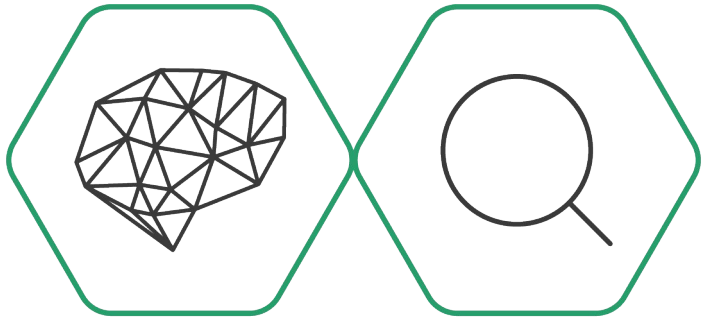
Apprenez à détecter les images falsifiées

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

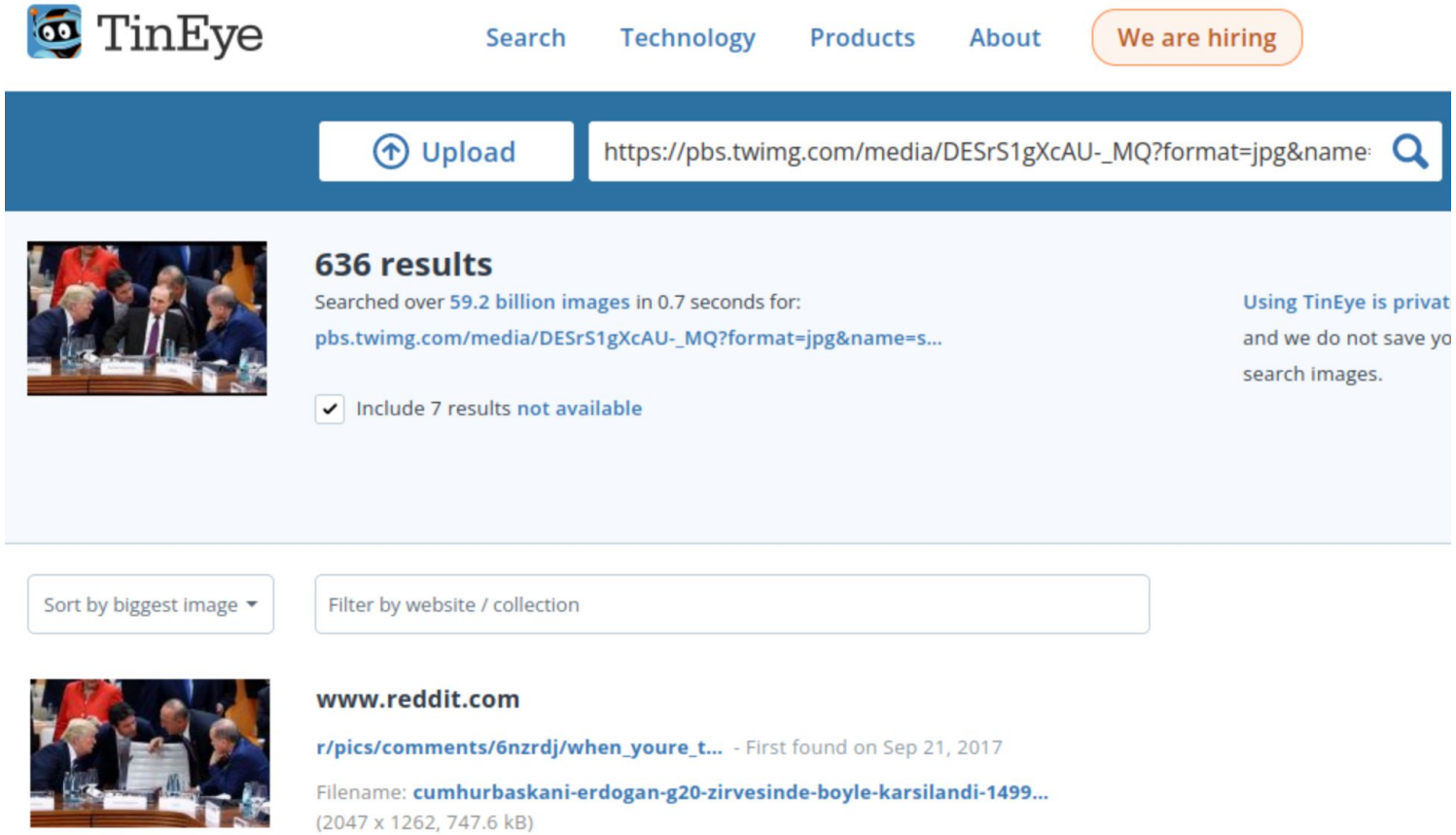
Les images falsifiées sont partout : fake news

Les images falsifiées sont partout : fake news

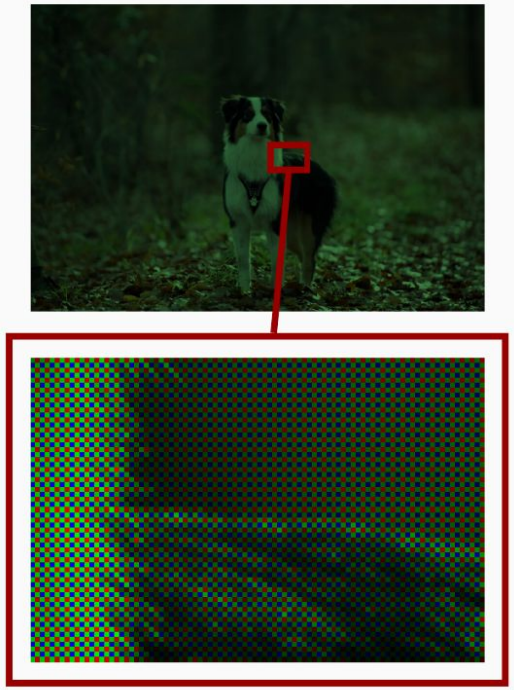
La photo a été prise à Londres, et a été falsifiée pour la détourner de son contexte !

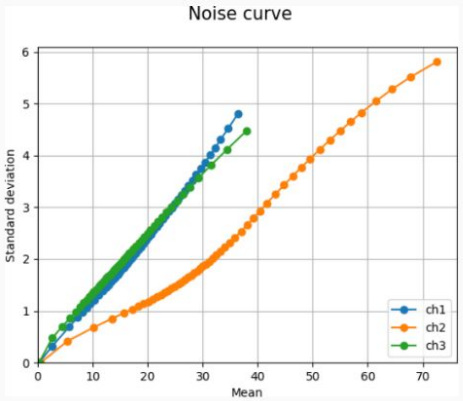
Comment savoir si une image est falsifiée ? Recherche inversée

- Google Images, TinEye

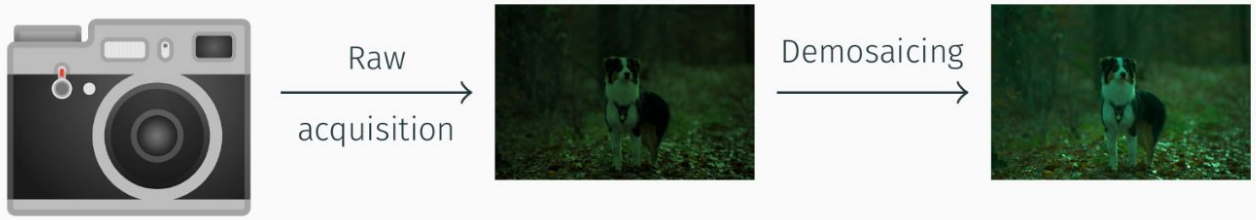


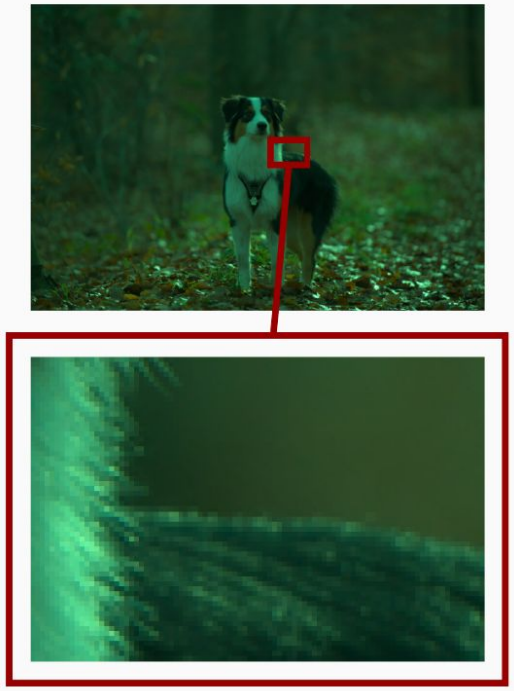
La vie secrète d'une image

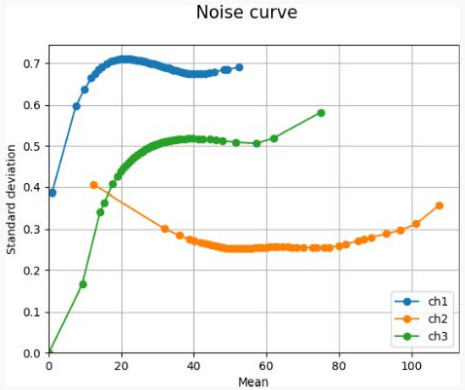




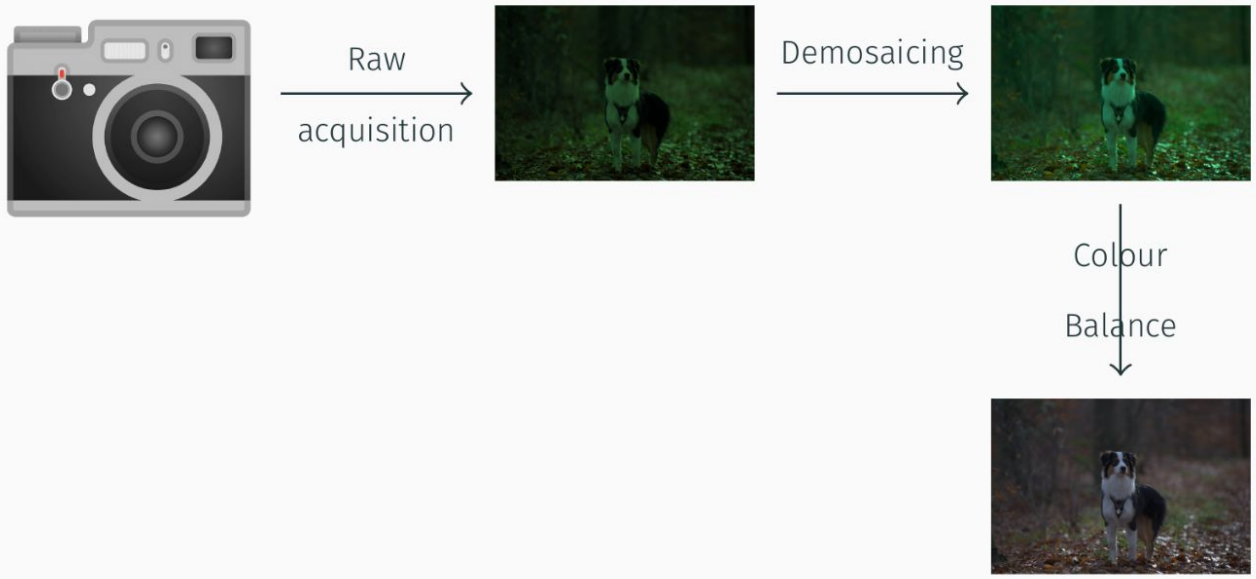
La vie secrète d'une image

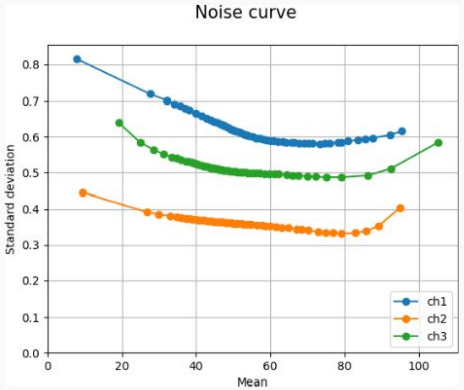




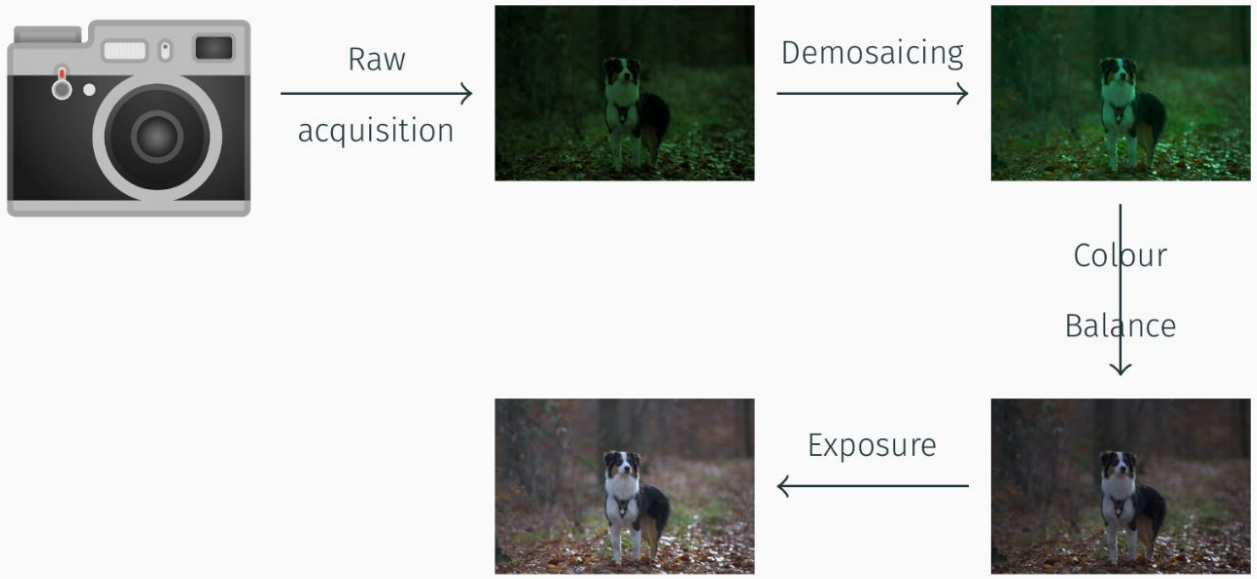


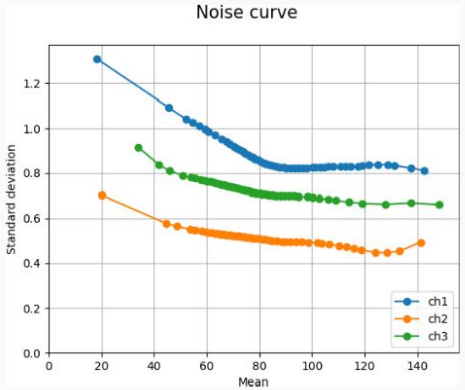
La vie secrète d'une image



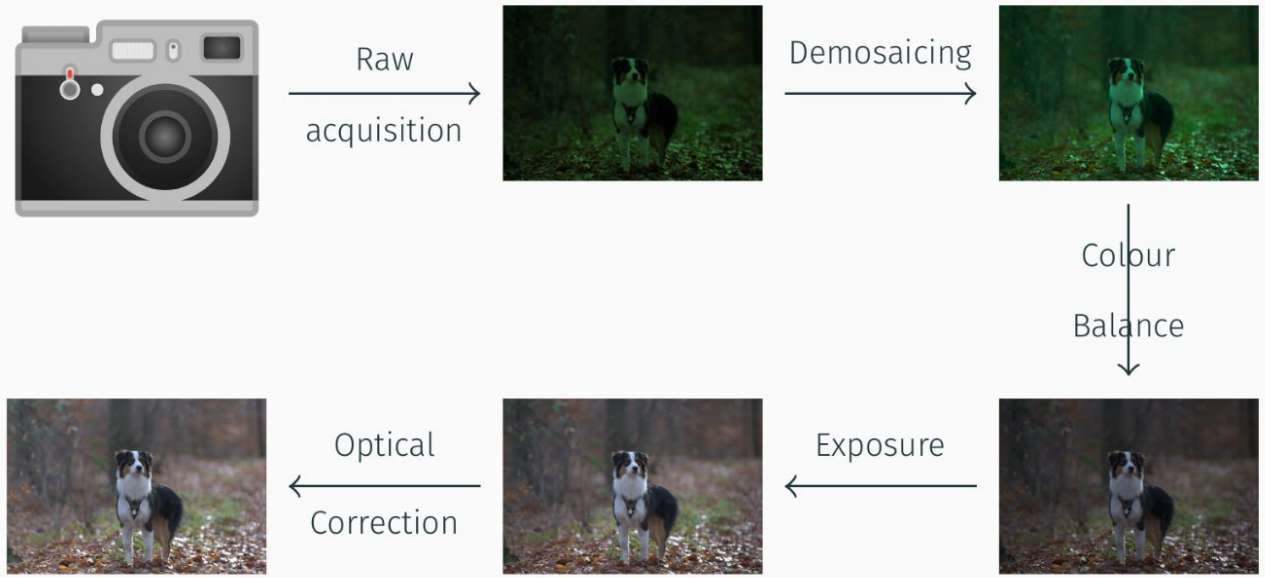


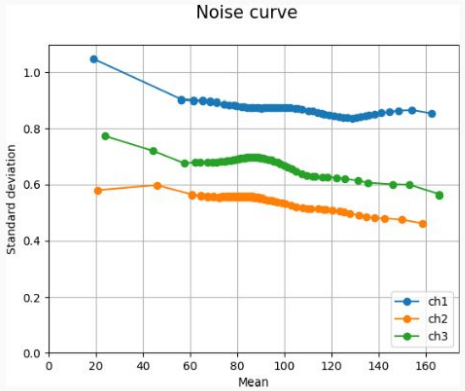
La vie secrète d'une image



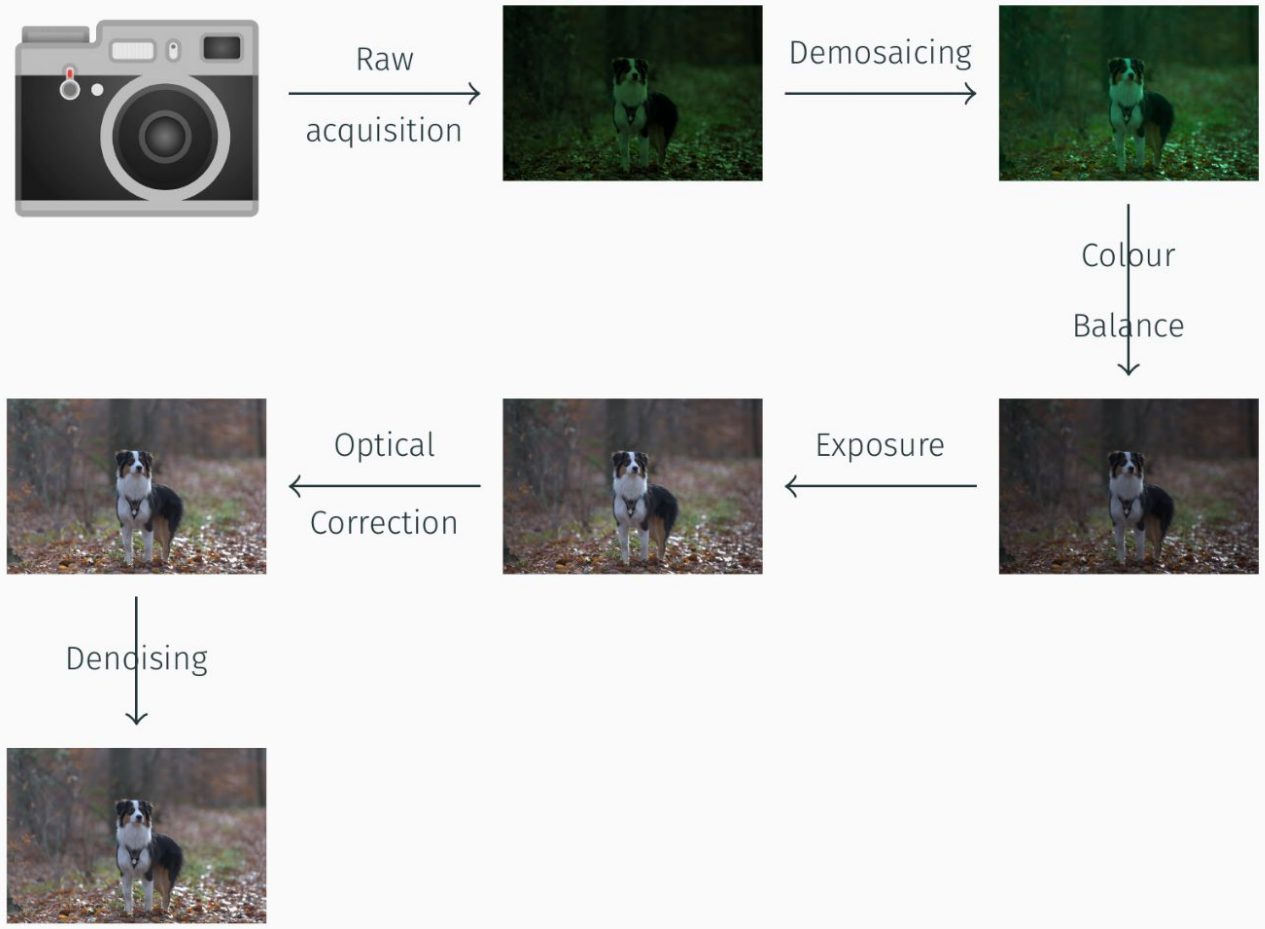


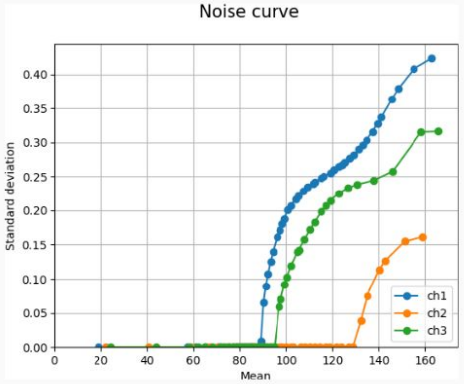
La vie secrète d'une image



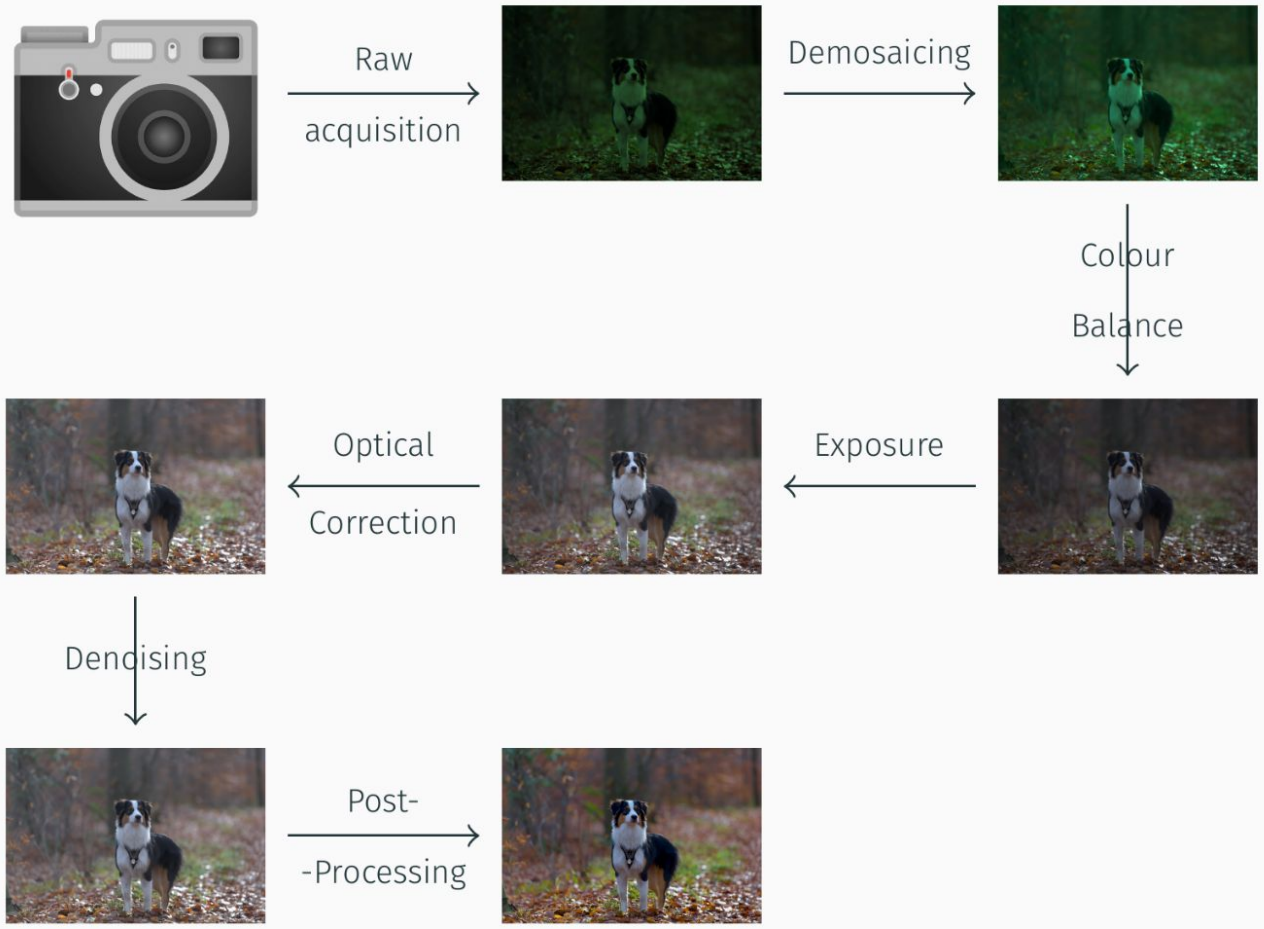


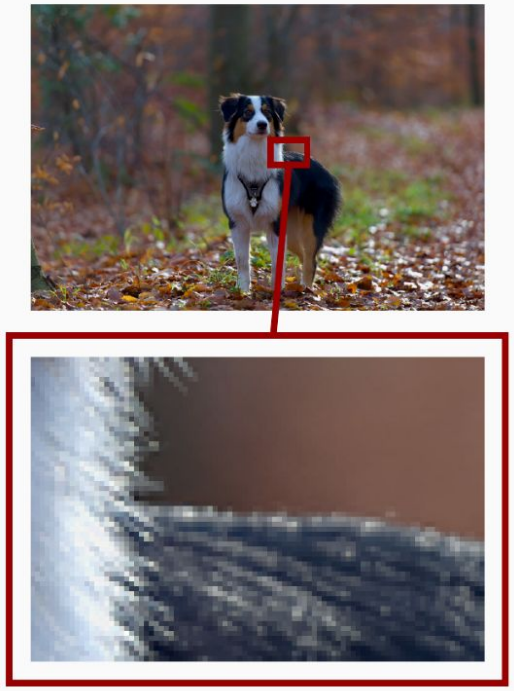
La vie secrète d'une image

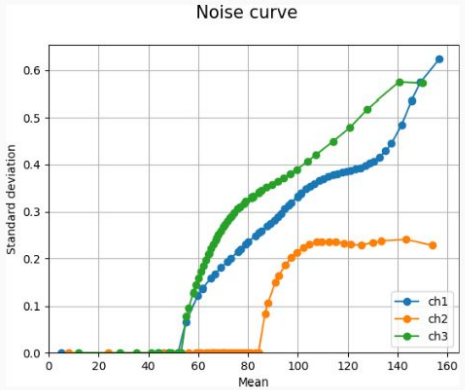




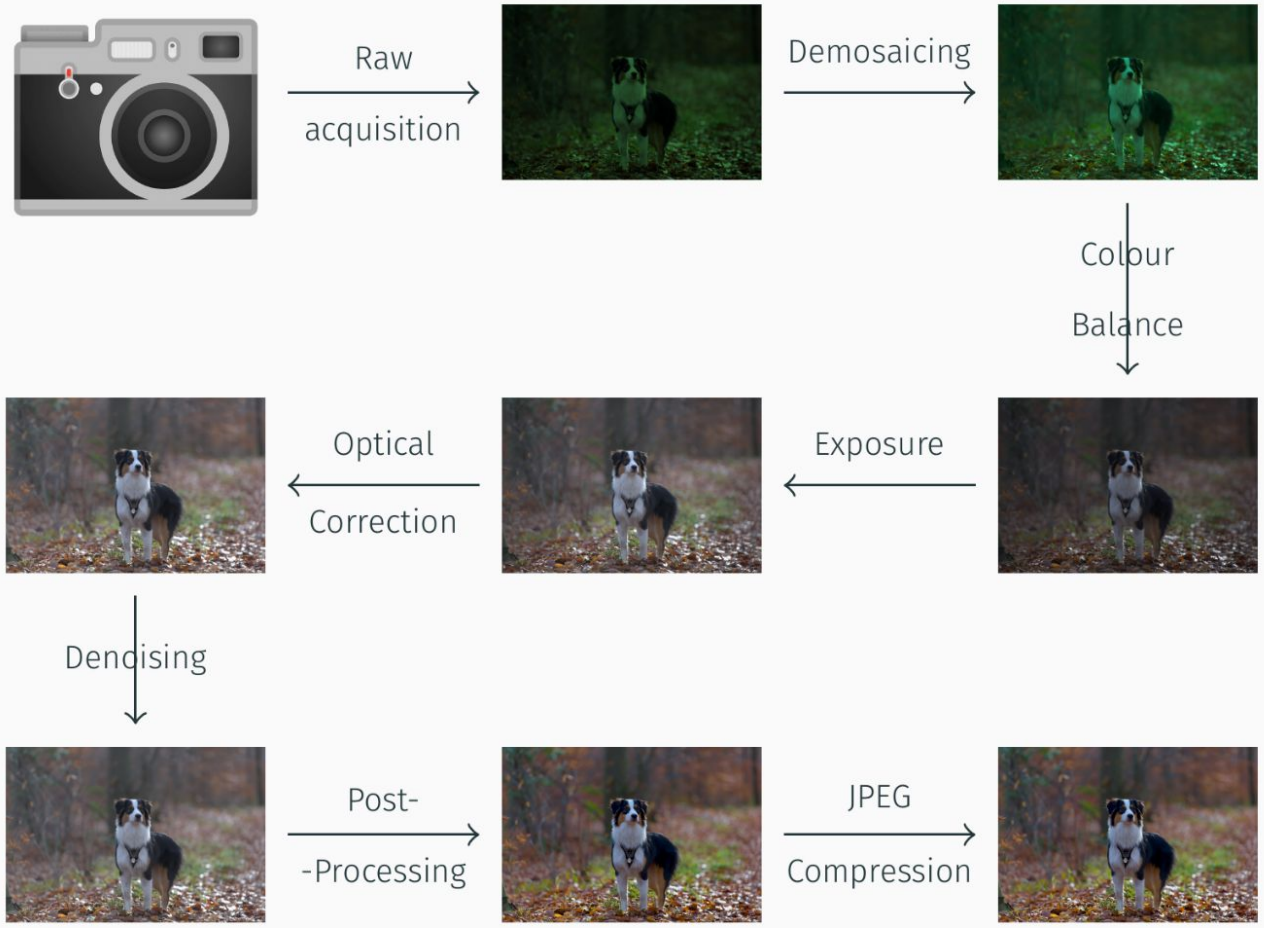
La vie secrète d'une image

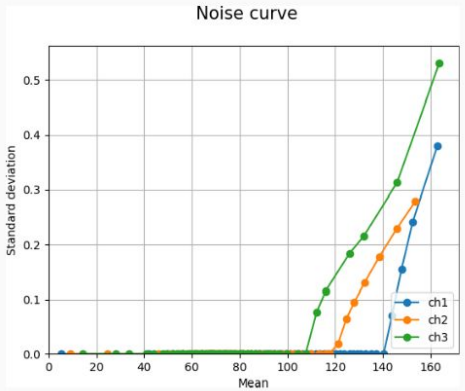




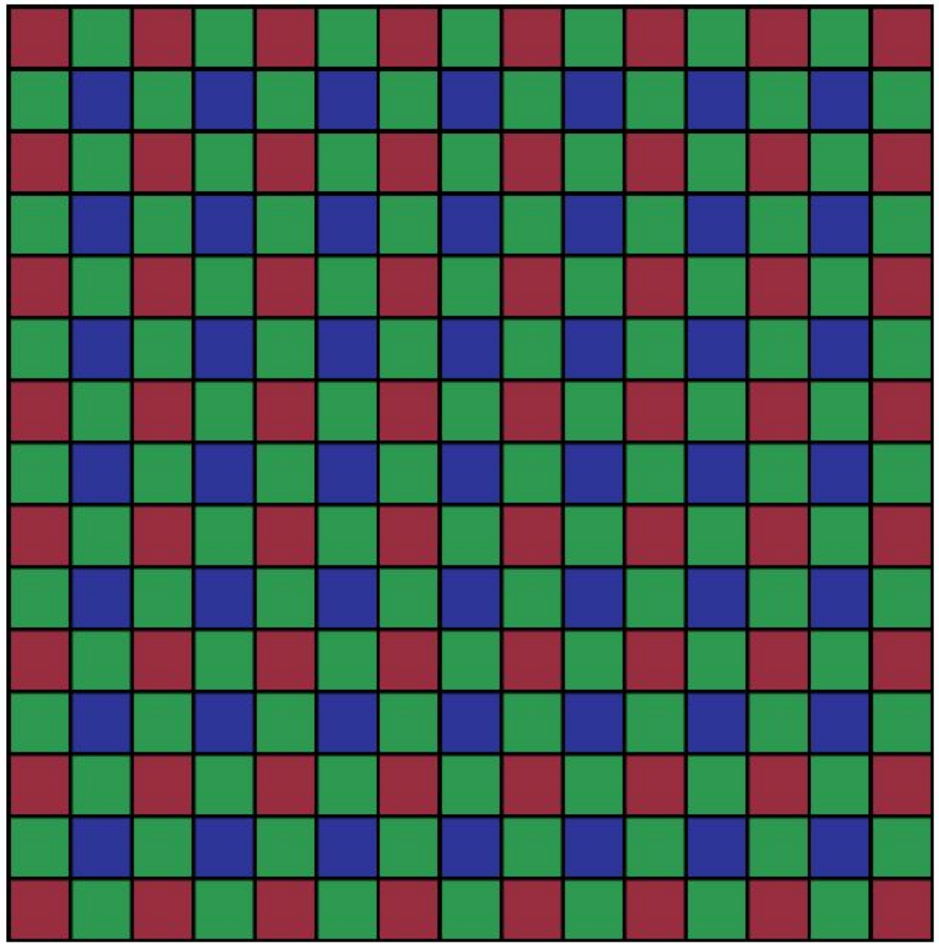


La vie secrète d'une image

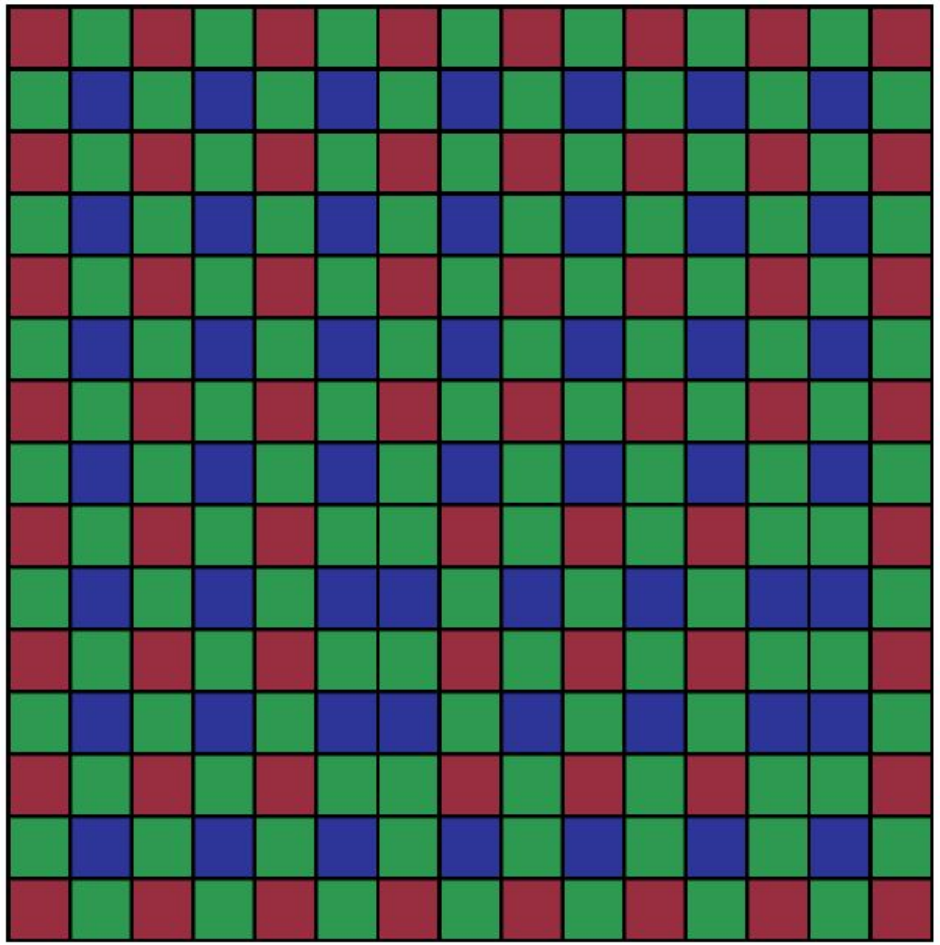




Détection de falsifications à travers l'analyse de démosaïquage

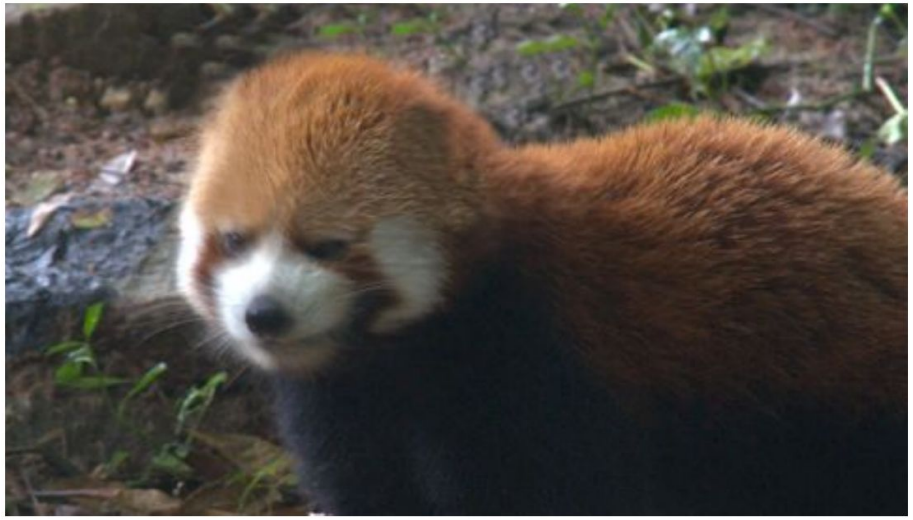


(a) Authentic image



(b) Forged image

Détection de falsifications à travers l'analyse de démosaïquage

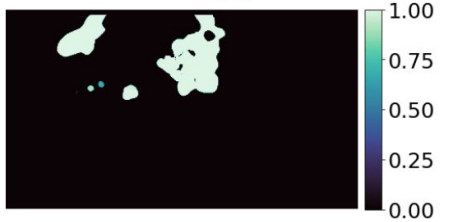


(a) Forged image

(b) Original image

(c) Forgery mask

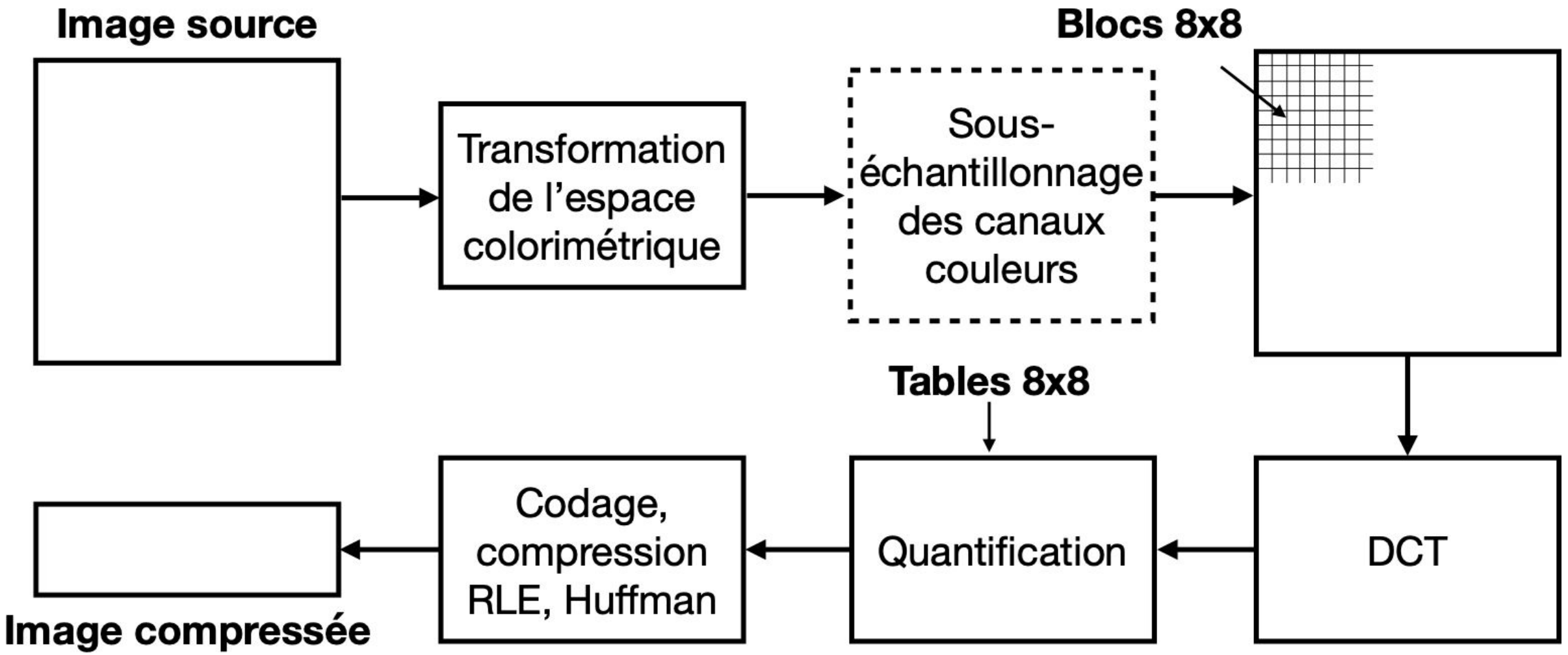
(d) Detected demosaicing grids



(e) Detected forgeries

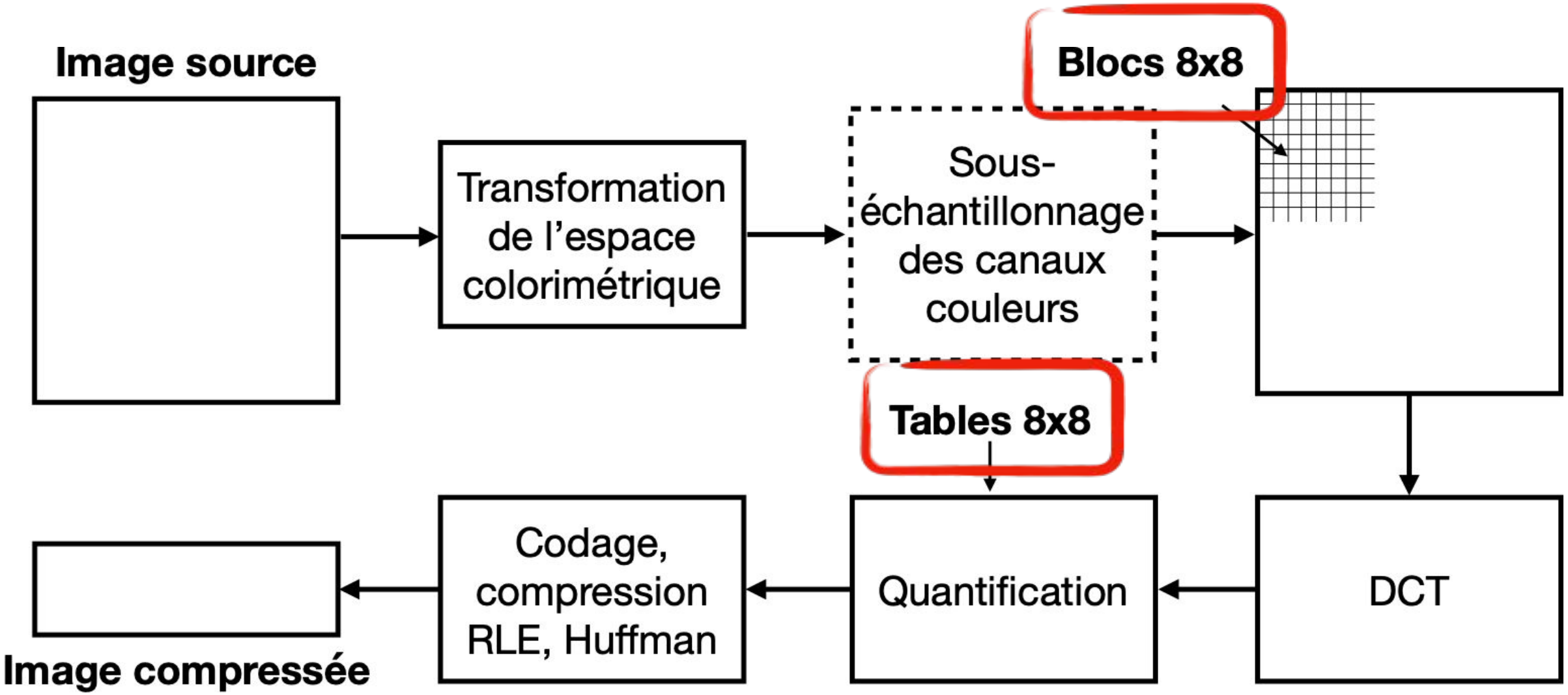
Source : base de données Korus

Détection de falsifications à travers l'analyse de compression JPEG



Crédit Tina Nikoukhah

Détection de falsifications à travers l'analyse de compression JPEG

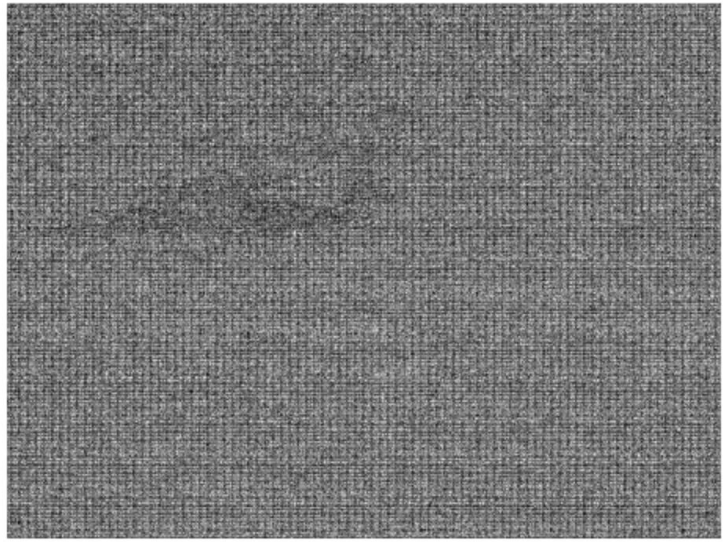


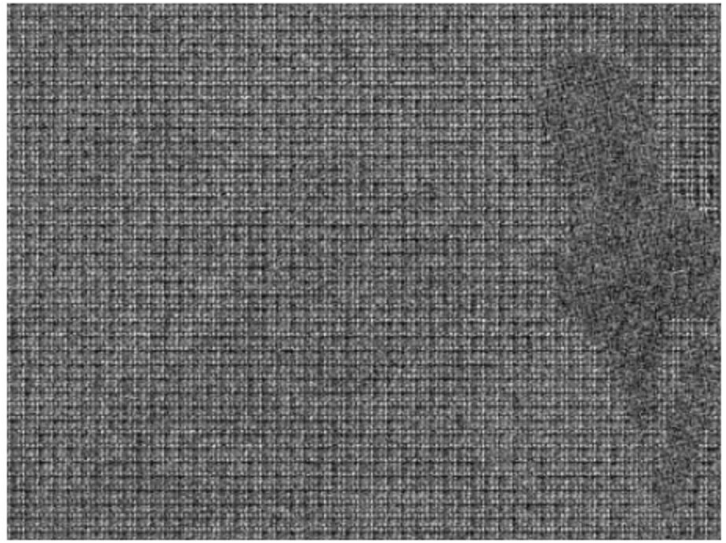
Crédit Tina Nikoukhah

Détection de falsifications à travers l'analyse de compression JPEG

Crédit Tina Nikoukhah

Détection de falsifications grâce aux textures



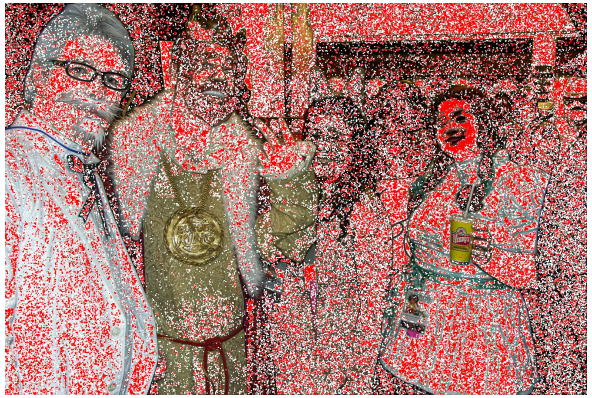


Cozzolino, D., & Verdoliva, L. (2019). Noiseprint: A CNN-based camera model fingerprint. *IEEE Transactions on Information Forensics and Security*, *15*, 144-159.

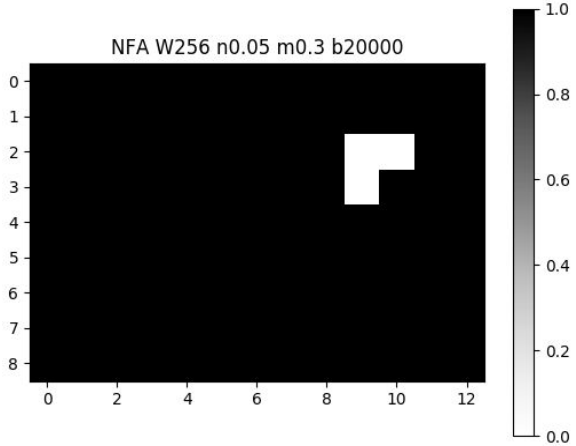
Détection de falsifications à travers l'analyse de bruit

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

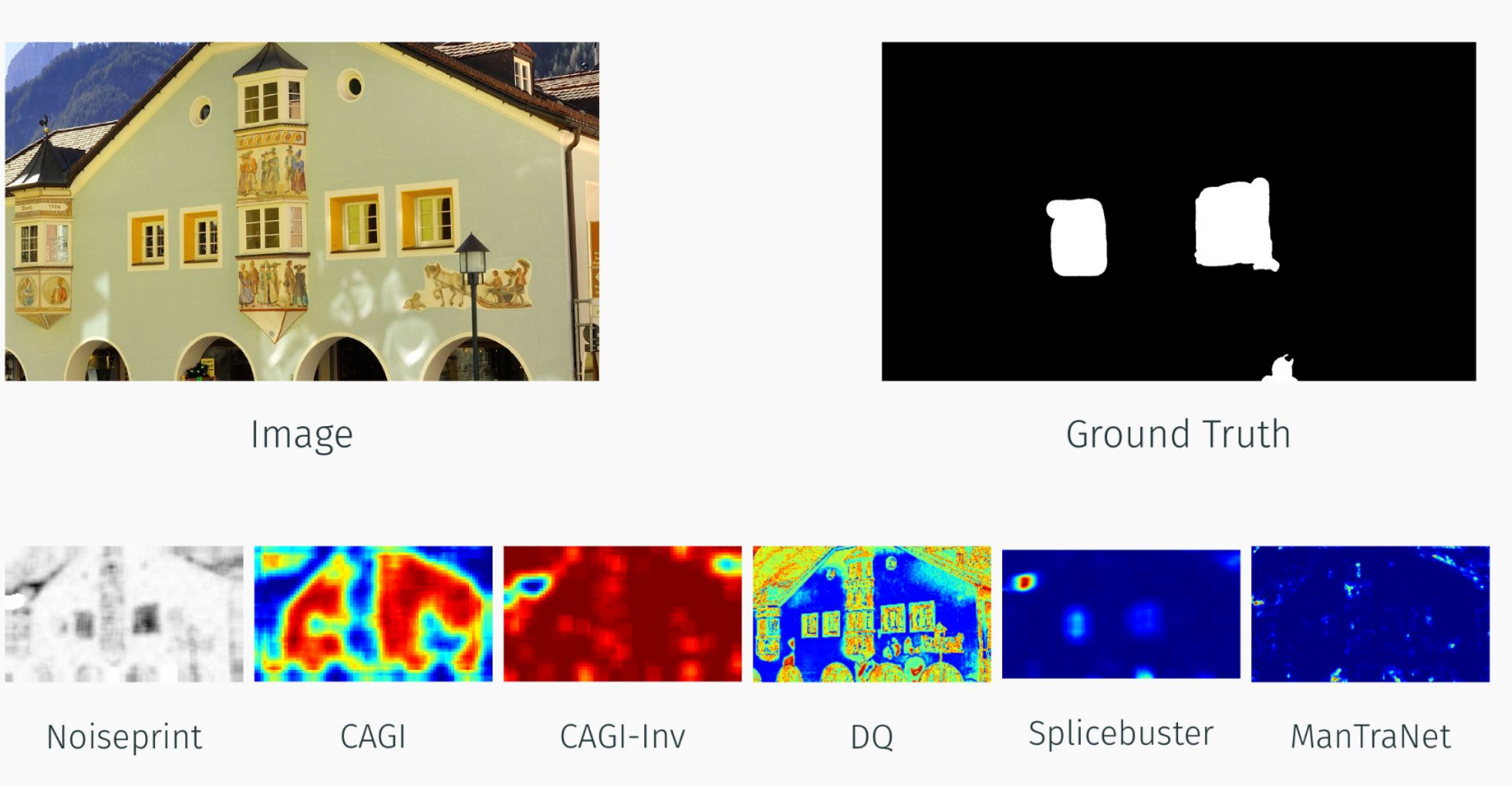
Figure 2: Deux descripteurs similaires (dont les positions sont représentés par un point rouge) d'une région clonée: Les descripteurs sont identiques.

Détection de copier-coller

Figure 3: Objets similaires mais différents (base COVERAGE)

Figure 4: Image falsifiée (base COVERAGE)

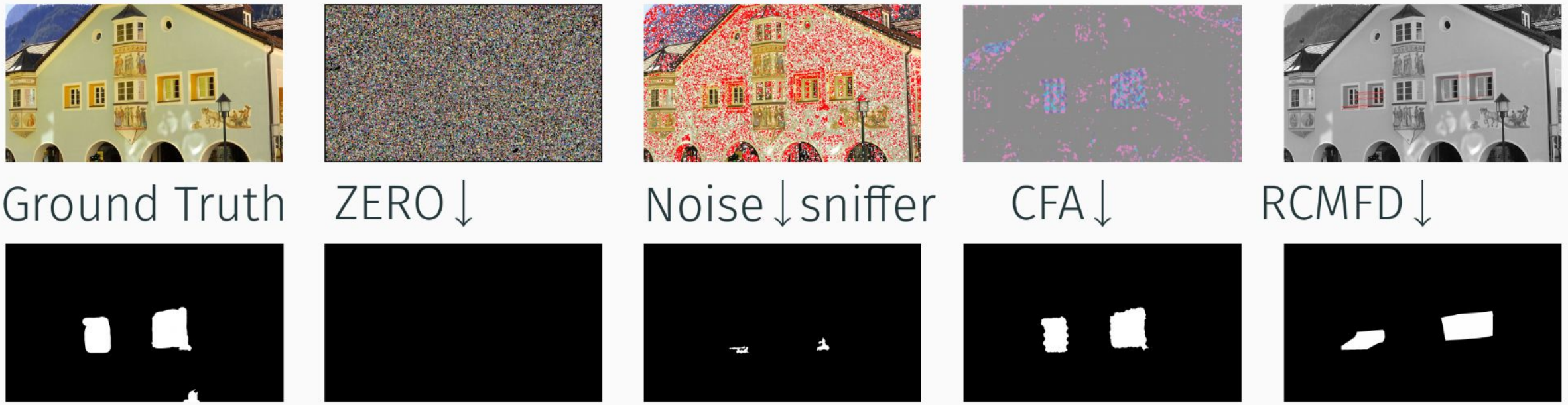
Comment détecter une falsification ?
Des méthodes obscures…



Beaucoup de méthodes existantes nécessitent d'être interprétées :

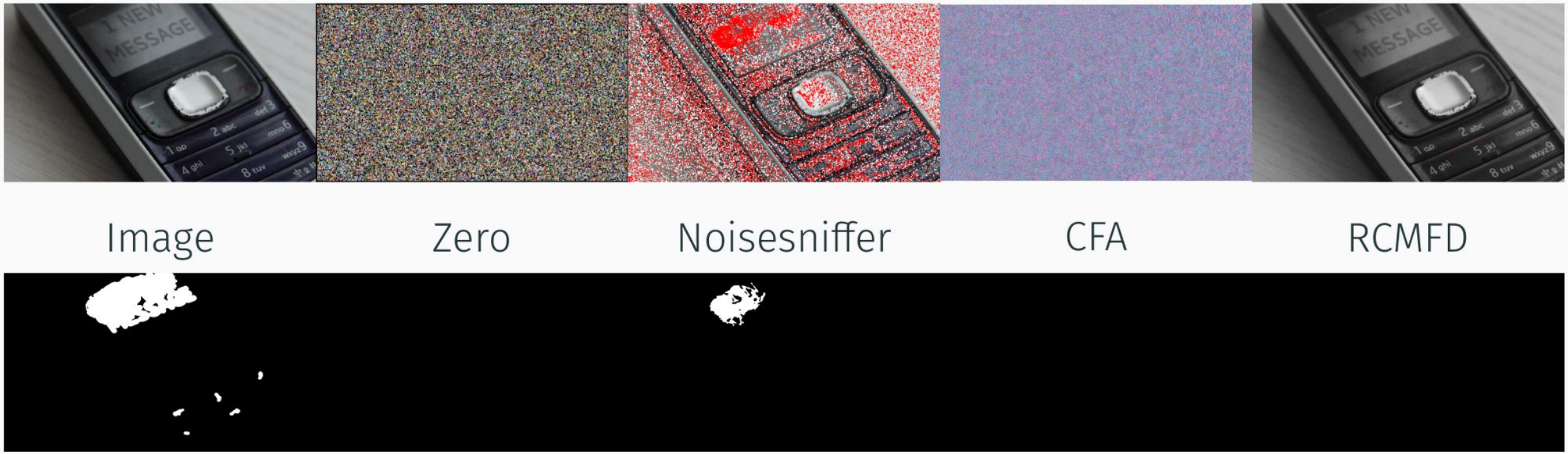
- Quelles parties correspondent vraiment à une falsification ?
- Quelles méthodes sont fiables ?
- Sur quelles justifications basent-elles leur détection ?

Comment détecter une falsification ?
… Vers des méthodes interprétables



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



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

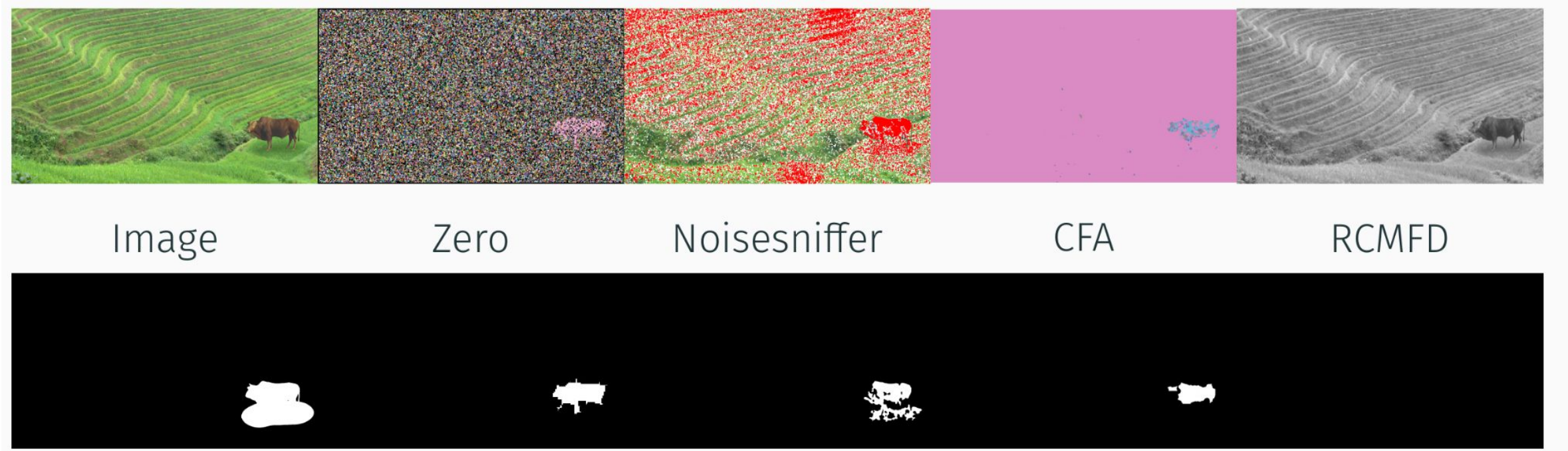
Ça ne veut pas dire que l'image est authentique !

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 :

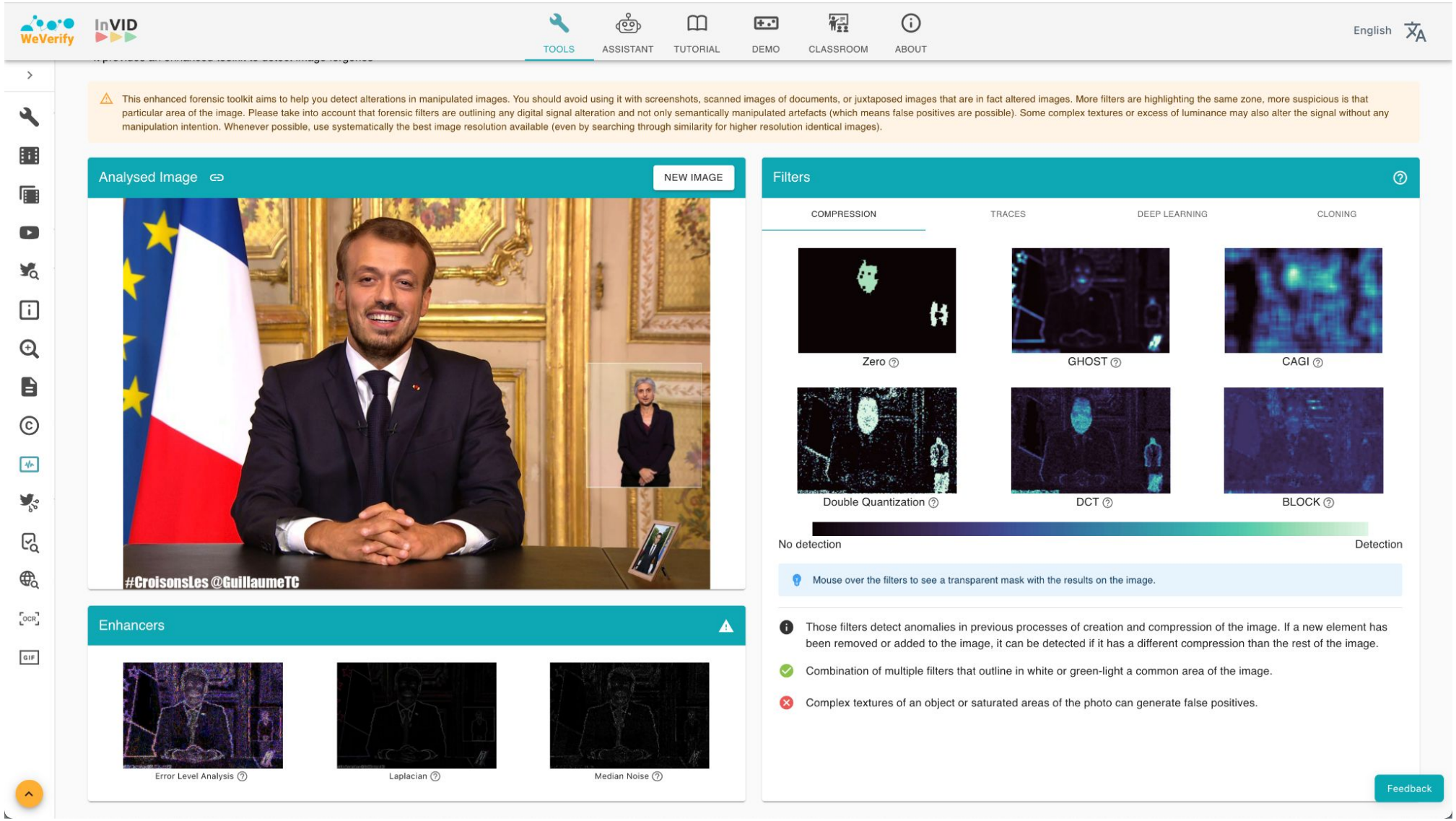
- RCMFD ne détecte rien = pas de copier-coller visible

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

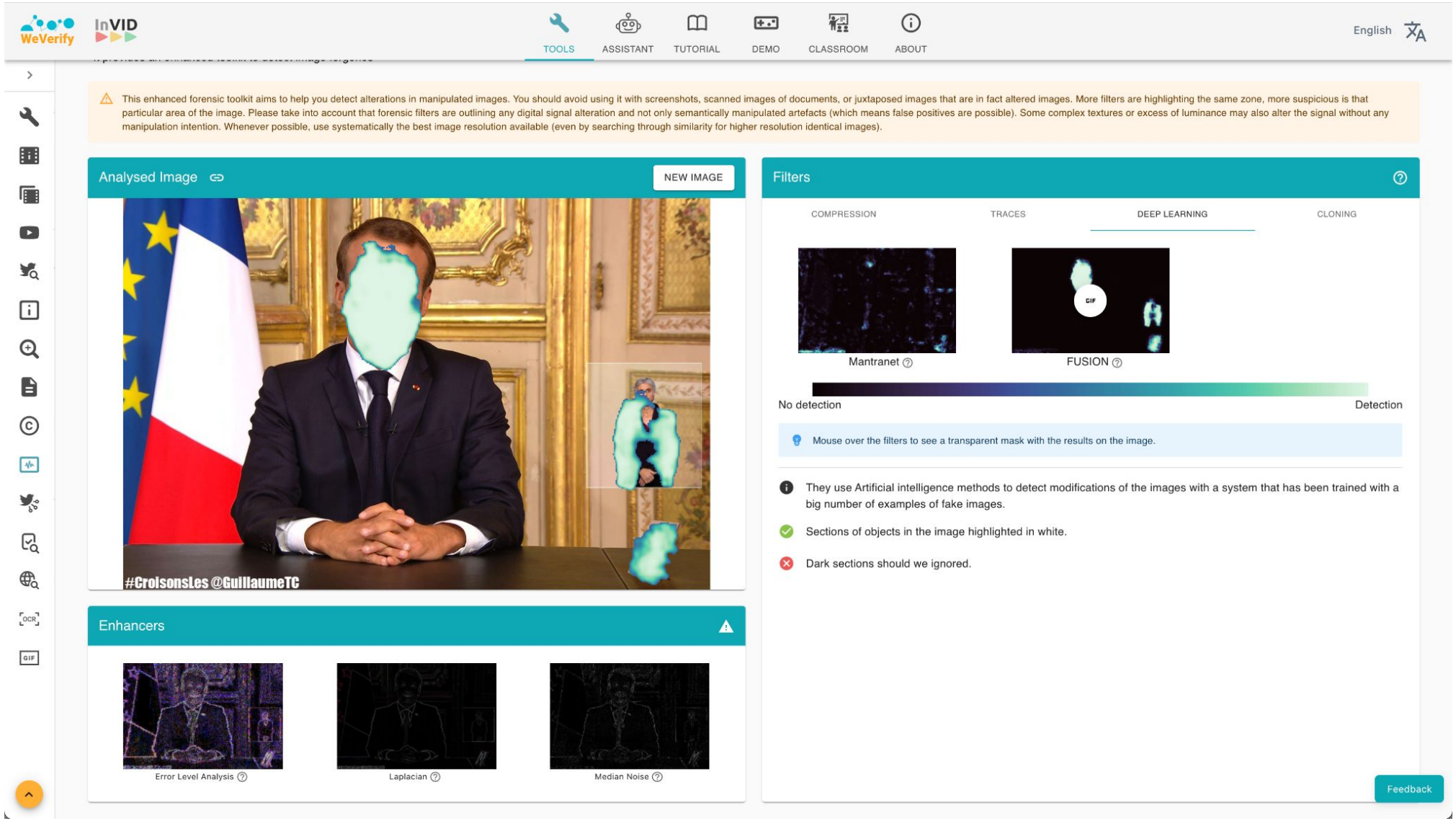
Comment utiliser des outils de détection de falsification ?

- 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
- Démos IPOL
  - IPOL : Journal scientifique en ligne
  - Chaque article est accompagné d'une démo utilisable par tous
  - Examples :
    - 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/

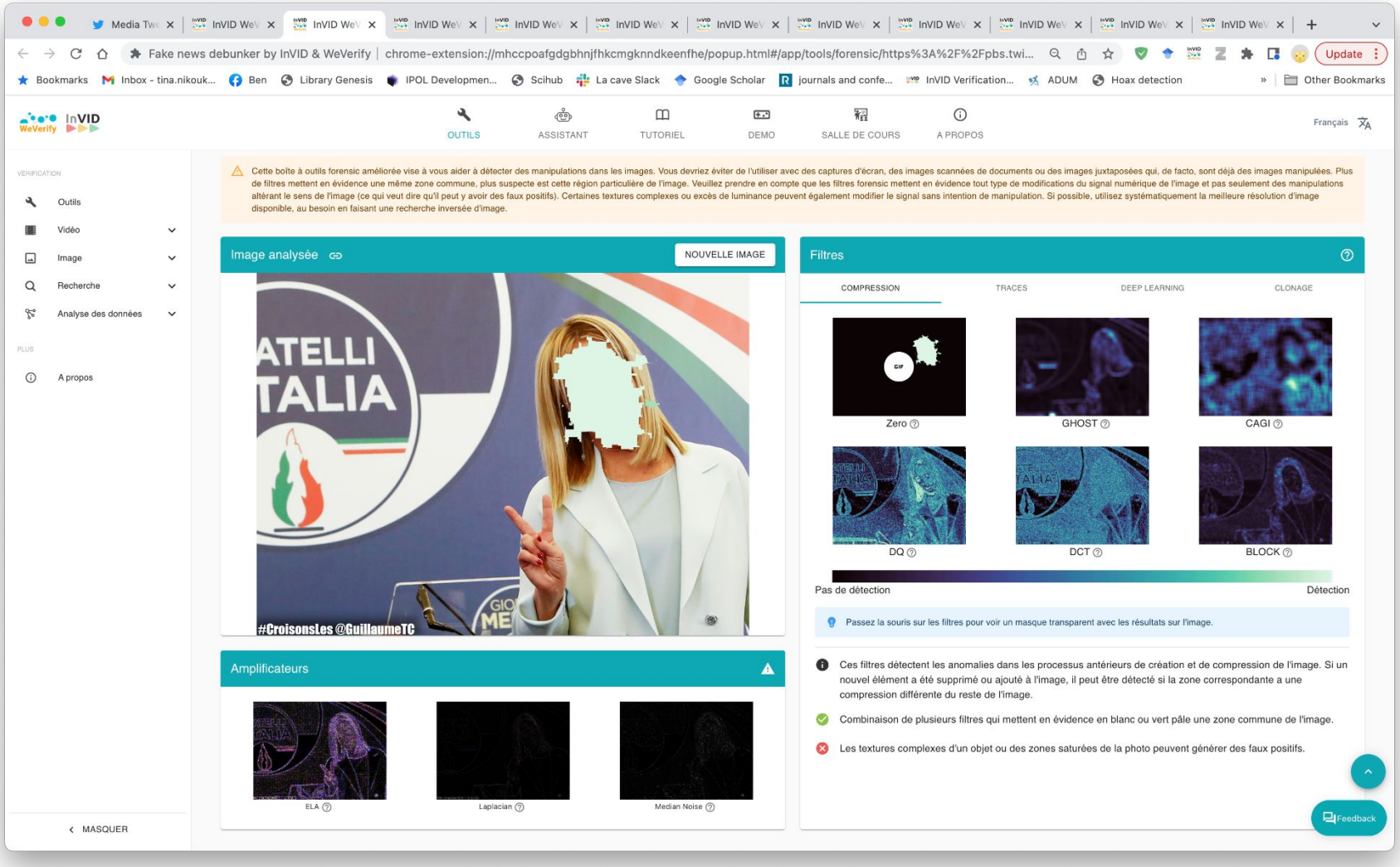
Comment utiliser des outils de détection de falsification ?



Comment utiliser des outils de détection de falsification ?



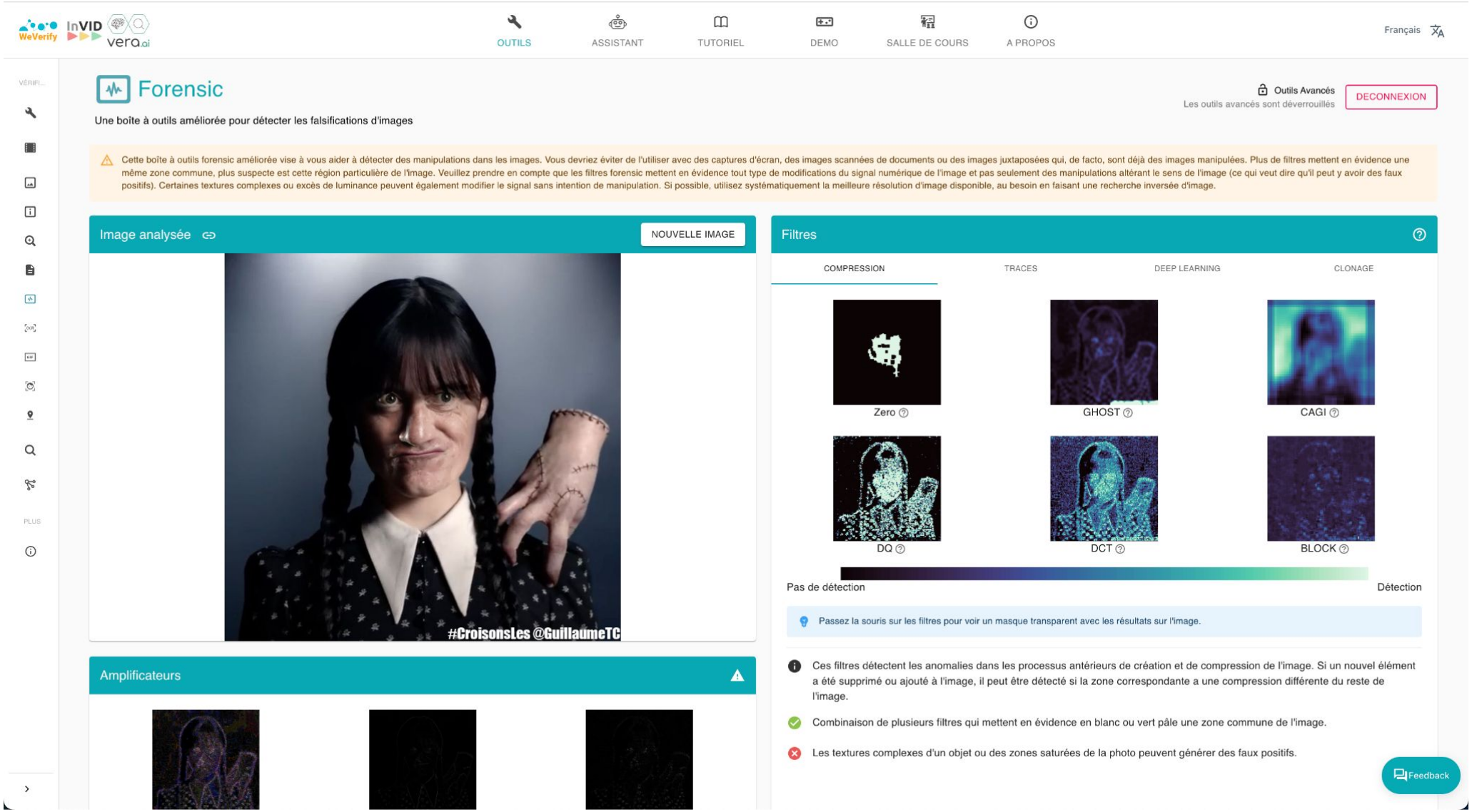
Comment utiliser des outils de détection de falsification ?



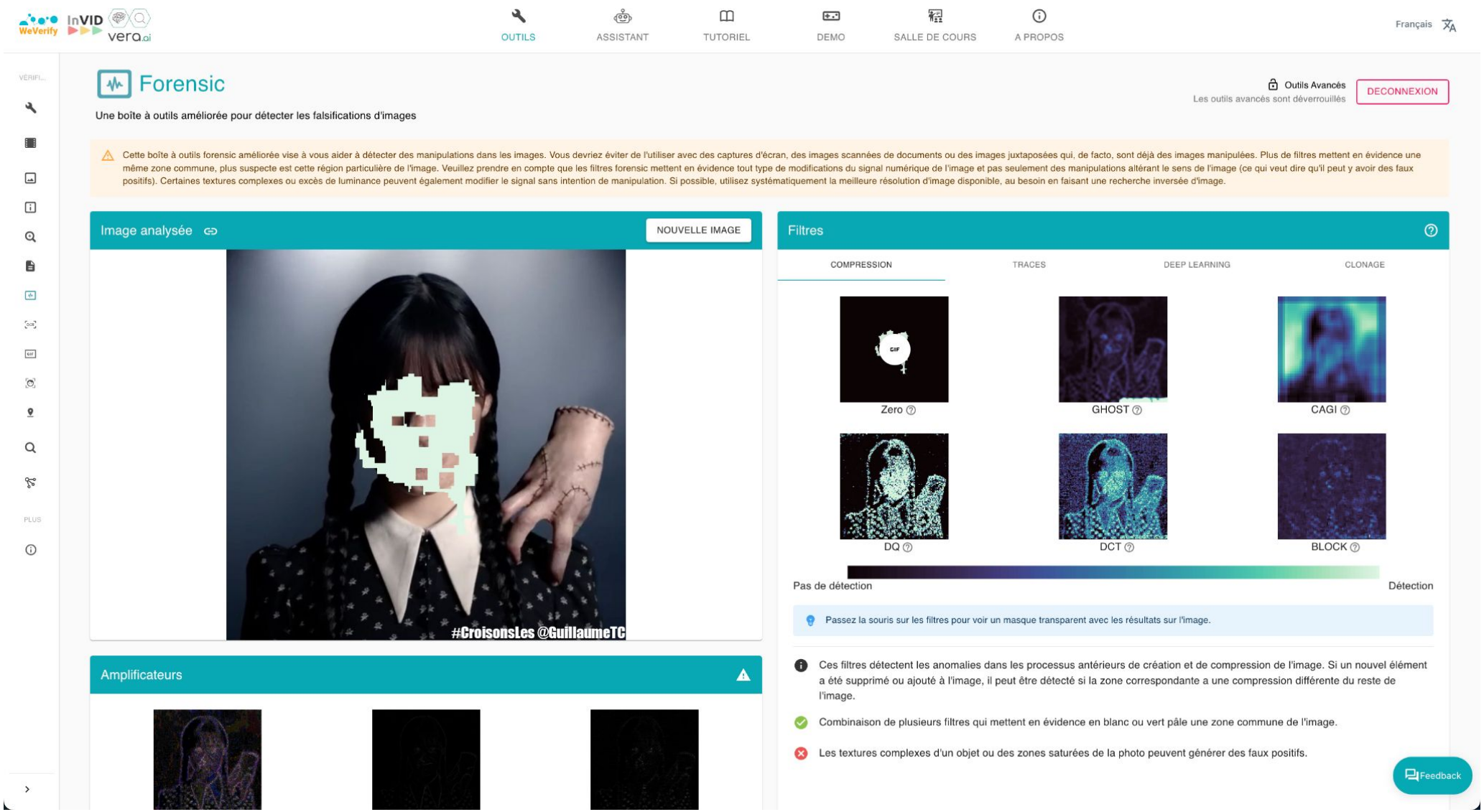
Comment utiliser des outils de détection de falsification ?

Comment utiliser des outils de détection de falsification ?

Comment utiliser des outils de détection de falsification ?



Comment utiliser des outils de détection de falsification ?



Quel futur pour les falsifications ?
Des deepfakes…

https://youtu.be/cQ54GDm1eL0

Quel futur pour les falsifications ?
… aux images de synthèse

vera.ai

Quel futur pour les falsifications ?
… aux images de synthèse

Quel futur pour les falsifications ?
… aux images de synthèse

vera.ai

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

vera.ai

CERTH CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

The University Of Sheffield.

ATC ATHENS TECHNOLOGY CENTER

KINIT

ontotext

Deutsche Welle

Università degli Studi di Napoli Federico II

1506 Università degli Studi di Urbino Carlo Bo

AFP

Universiteit van Amsterdam

école normale supérieure paris-saclay

université PARIS-SACLAY

EBU OPERATING EUROVISION AND EURORADIO

EU DisinfoLab

Fraunhofer IDMT

Follow us on Twitter: @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