

Titre: Détection de falsification d'images via l'analyse du démosaïquage : dévoilement d'une signature

Mots clés: Forensiques d'image, apprentissage positionnel, apprentissage auto-supervisé, apprentissage interne, démosaïquage

Résumé: Autrefois considérées comme des preuves fiables, les images photographiques ne dépeignent plus toujours la pure vérité. Avec l'avènement de la photographie numérique et les progrès des outils de retouche photo, il n'a jamais été aussi facile de modifier une image. Si la plupart de ces modifications visent uniquement à améliorer l'image, elles peuvent potentiellement en altérer la sémantique même. Dissimuler, modifier ou ajouter un objet étranger, tout cela peut donner à une image un sens nouveau et trompeur. Bien que ces falsifications puissent facilement être rendues visuellement réalistes, ils n'en altèrent pas moins le tissu même de l'image. La formation d'une image numérique, depuis les capteurs de la caméra jusqu'au stockage, laisse des traces, qui agissent comme une signature de l'image. La modification d'une image déforme ces traces, créant des incohérences détectables.

Les images brutes sont initialement une mosaïque de pixels rouges, bleus et verts. Les valeurs de couleur manquantes doivent être interpolées dans un processus connu sous le nom de démosaïquage. Dans cette thèse, nous étudions les traces laissées par ce processus. La nature 2-périodique du motif de la mosaïque laisse son empreinte sur l'image. Les falsifications peuvent déphaser ces traces, voire les supprimer entièrement ; l'identification du motif de mosaïque est donc utile pour localiser les régions falsifiées.

Les méthodes non spécifiques de détection des falsifications peuvent déjà analyser de nombreuses traces dans une image ; elles restent néanmoins aveugles aux déplacements de la mosaïque, en raison de l'invariance par translation des réseaux de neurones convolutifs sur lesquels la plupart sont basés. Les méthodes spécifiques au démosaïquage peuvent donc fournir des résultats complémentaires pour la détection des falsifications. Cependant, elles ont historiquement reçu peu d'attention. L'analyse des artefacts de démosaïquage est rendue plus difficile par la vaste gamme d'algorithmes

de démosaïquage, souvent non divulgués, et surtout par la compression JPEG. Ces artefacts, créés tôt dans le pipeline de formation de l'image et situés aux fréquences les plus élevées de l'image, s'estompent rapidement pendant la compression.

Pourtant, ces artefacts peuvent encore être détectés sous une compression légère. Pour canaliser la puissance représentative des réseaux neuronaux convolutifs dans l'analyse des artefacts de démosaïquage, nous introduisons la notion d'apprentissage positionnel. Ce schéma auto-supervisé entraîne le réseau à détecter la position modulo 2 de chaque pixel, en tirant parti de l'invariance de translation de la convolution pour que le réseau analyse implicitement les artefacts de démosaïquage, son seul indice de la position modulo 2 d'un pixel. De plus, l'entraînement interne sur une seule image potentiellement falsifiée peut renforcer la robustesse de la méthode à la compression JPEG de ladite image. Les erreurs dans la sortie du réseau neuronal sont alors des indices d'incohérences de la mosaïque. Un paradigme *a contrario* nous permet alors de prendre des décisions automatiques sur l'authenticité d'une image. En utilisant uniquement les artefacts de démosaïquage, la méthode proposée dépasse l'état de l'art sur plusieurs jeux de données non compressés. Sur les images compressées, elle fournit encore des résultats décents qui sont tout à fait complémentaires avec les méthodes qui ne sont pas spécifiques à la mosaïque.

Enfin, nous explorons l'évaluation même des méthodes de détection de falsification. Nous proposons une méthodologie et un jeu de données pour étudier la sensibilité des outils de détection à des traces spécifiques, ainsi que leur capacité à effectuer des détections sans indices sémantiques sur l'image. Plus qu'un simple outil d'évaluation, cette méthodologie peut être utilisée pour évaluer les forces et faiblesses de chaque méthode, ainsi que leurs complémentarités.

Title: Image Forgery Detection through Demosaicing Analysis: Unconcealment of a Signature

Keywords: Image Forensics, Positional Learning, Self-Supervised Learning, Internal Learning, Demosaicing

Abstract: Once considered reliable evidence, photographic images can no longer be assumed to depict the naked truth. With the advent of digital photography and the progress of photo editing tools, altering a picture has never been easier. While most of these modifications solely seek to enhance the image, they can potentially alter its very semantics. Concealing, modifying or adding a foreign object, all those can give an image a new and false meaning. Although these forgeries can easily be made visually realistic, they still distort the very fabric of the image. The formation of a digital image, from the camera sensors to storage, leaves traces, which act like a signature for the image. Modifying an image distorts these traces, creating detectable inconsistencies.

Raw images are initially a mosaic of red, blue and green pixels. Missing colour values must be interpolated in a process known as demosaicing. In this thesis, we study the traces left by this process. The 2-periodic nature of the mosaic pattern leaves its imprint onto the image. Forgeries may dephase these traces, or even remove them entirely; mosaic pattern identification is consequently helpful in localizing tampered regions.

Non-specific forgery detection methods can already analyse many traces in an image; nevertheless they remain blind to shifts in the mosaic, due to the translation-invariance of the convolutional neural networks on which most are based. Demosaicing-specific methods can thus provide complementary results for forgery detections. However, these have historically received little attention. Analysis of demosaicing artefacts is made harder by the vast array of often-undisclosed de-

mosaicing algorithms, and above all by JPEG compression. Those artefacts, created early in the image formation pipeline and lying at the highest frequencies of the image, are quick to wane during compression.

Yet, those artefacts can still be detected under mild compression. To channel the representative power of convolutional neural networks into the analysis of demosaicing artefacts, we introduce the notion of positional training. This self-supervised scheme trains the network to detect the modulo-2 position of each pixel, leveraging the translation invariance of convolution to make the network implicitly analyse demosaicing artefacts, its only clue to the modulo-2 position of a pixel. On top of that, internal training on a single potentially forged image can bolster the method's robustness to JPEG compression on said image. Errors in the output of the neural network are then clues of mosaic inconsistencies. An *a contrario* paradigm then enables us to make automatic decisions on the authenticity of an image. Using only demosaicing artefacts, the proposed method beats the state of the art on several uncompressed datasets. On compressed images, it still provides decent results that are fully complementary with methods that are not mosaic-specific.

Finally, we explore the very evaluation of forgery detection methods. We propose a methodology and dataset to study the sensitivity of forensic tools to specific traces, as well as their ability to make detections without semantic cues on the image. More than a simple evaluation tool, this methodology can be used to assess the strength and weaknesses of each method, as well as their complementarities.