Etude Comparative de l'Apprentissage par Transfert pour l'Identification des Caméras

Alexandre Berthet, Jean-Luc Dugelay
Eurecom, Département Sécurité Digitale, Sophia Antipolis, France.

Résumé : *De nos jours, l'édition d'images est devenue plus facile et plus précise. Les modifications malveillantes sont donc plus accessibles et des méthodes de détection des falsifications ont été développées. L'identification de la caméra source est un domaine important de la criminalistique des images numériques et peut-être réalisée selon la marque, le modèle ou l'appareil exact. Notamment, la classification de modèle de caméras est l'application la plus abordée et a d'abord été étudiée avec des algorithmes classiques, puis avec des réseaux de neurones convolutifs (RNC). Cependant, malgré leur efficacité, les RNC sont dépendants des données et leurs performances diminuent avec le nombre d'objets à classer. Cet aspect est encore plus important avec les caméras puisque chaque appareil possède ses propres artefacts et que les caméras d'une même marque présentent des similarités au niveau de leurs empreintes numériques. Dès lors, un des objectifs de cet article réside dans l'étude de la robustesse des méthodes pour l'identification de caméras. Un domaine récent de l'apprentissage automatique, appelé apprentissage par transfert, offre une alternative intéressante à ce problème. Afin d'étudier pleinement son impact, nous avons appliqué différentes approches de l'apprentissage par transfert à trois architectures de RNC différentes, présentant chacune des particularités. Pour réaliser notre étude comparative, nous avons également proposé un protocole d'évaluation de la robustesse basé sur les deux principales sujet de recherche de l'état de l'art : les caméras inconnues et l'augmentation de caméras à classer.*

Mots-clés : Identification du modèle de caméra ; Apprentissage par transfert ; Criminalistique des images numériques ; Réseaux de neurones convolutifs.

1 Introduction

Grâce au développement des dispositifs numériques (appareils photo, téléphones portables, etc.), l'accès aux images et aux vidéos a augmenté au point de devenir un important canal de communication. D'autre part, la retouche des images numériques est devenue un véritable problème, notamment pour prouver l'authenticité d'une image. Dans le même temps, l'identification de la caméra source, un domaine de la criminalistique des images numériques [11], s'est révélée être une solution pour la détection de ces falsifications. Cette identification peut être faite selon la marque, le modèle ou le dispositif numérique même de la caméra. La classification de modèle de caméras est l'application la plus étudiée dans la littérature et le principe est d'identifier, parmi un groupe de modèles de caméras, celui qui est associé à l'image étudiée. Plusieurs techniques ont été élaborées, utilisant les artefacts laissés lors de l'acquisition d'une image numérique. Notamment, le bruit numérique lié au capteur de la caméra [10] ou à des caractéristiques physiques [6] a d'abord permis d'établir l'empreinte numérique des caméras. Puis, avec la démocratisation de l'apprentissage automatique, les performances ont été améliorées, notamment grâce aux réseaux de neurones convolutifs (RNC). Ces réseaux analysent d'abord les artefacts et les réduisent via un extracteur de caractéristiques, tandis que l'identification est réalisée grâce aux couches de classification. Cependant, même si ces méthodes se sont avérées efficaces pour identifier les modèles de caméras, il existe toujours une forte dépendance aux données. Ce phénomène est plus accentué pour les modèles de caméras puisque chaque dispositif numérique possède ses propres artefacts et que les caméras d'une même marque ont des empreintes numériques proches. Ces phénomènes d'unicité des artefacts et de similarité d'empreinte numérique posent la question de la robustesse de performance des méthodes. Dans cet article, nous avons utilisé l'apprentissage par transfert pour conduire une étude comparative de la robustesse des méthodes d'identification de modèle de caméras. De plus, cette étude est menée avec trois architectures de RNC réputées ainsi que les deux sujets de recherche principaux de l'identification de la caméra source.

La section 1 présente l'identification de la caméra source ainsi que les contributions de cet article. Dans la section 2, nous détaillons les motivations de notre étude comparative, les architectures utilisées ainsi que notre protocole. Les résultats de l'évaluation conduite à l'aide de notre protocole sont détaillés dans la section 3 ainsi que la base de données d'images de Dresde utilisée. Enfin, la section 4 conclut sur l'impact de l'apprentissage par transfert pour l'identification de modèle de caméras.

2 Travail proposé

2.1 Problème adressé

Le problème de similarité des empreintes numériques de modèles de caméras provenant d'une même marque est un des sujets importants de la littérature de l'identification de la caméra source. Dans [12], le problème est abordé à travers une série de trois expériences. La méthode est basée sur un RNC avec un filtre passe-haut utilisé comme module de prétraitement. Le prétraitement est une étape cruciale, voire obligatoire en criminalistique des images numériques [4]. i) Le réseau a tout d'abord été évalué avec 12 caméras de la base de données d'images de Dresde [7] (détaillée en section 3.1). ii) Ensuite, avec deux caméras supplémentaires provenant de la même marque

pour mettre en exergue le problème de similarité des empreintes numériques. iii) Enfin, avec l'ensemble des caméras (33 modèles) pour généraliser ce phénomène. Dans [5], le problème de similarité des modèles provenant d'une même marque est aussi adressé avec la classification de l'ensemble des modèles (27 caméras) de la base de données d'images de Dresde [7]. La méthode est basée sur un RNC pour l'extraction de caractéristiques et une machine à vecteurs de support pour la classification. En outre, le problème de caméras inconnues (c-à-d, non sélectionnées pour l'entraînement du réseau) est également abordé lors d'une expérience. Le problème des caméras inconnues est un sujet important de la littérature de l'identification des caméras. [2] est une étude focalisée sur le scénario de caméras inconnues avec une méthode basée sur un RNC dont la première couche est utilisée comme module de prétraitement. En fait, l'objectif de cette approche est de classer les images comme venant d'un modèle connu ou inconnu.

Le phénomène d'unicité des artefacts de dispositif numérique pose la question de la robustesse de performance des méthodes. Par exemple, une méthode ayant des performances élevées lors d'une évaluation sur une base de données *B1* (ayant servie pour l'entraînement) pourrait subir une baisse de performance sur une nouvelle base de données *B2*. En effet, si les caméras de la base *B2* sont inconnues du réseau, ce dernier pourrait ne pas classer correctement les modèles de caméras. Ces dernières années, l'apprentissage par transfert, un domaine de l'apprentissage profond, a été utilisé pour développer de nouveaux réseaux, de manière plus rapide sans perdre en efficacité. Le principe est de créer un nouveau réseau *M2* en transférant l'architecture et les poids d'un modèle *M1* pré-entraîné sur une base de données *B1*. Le réglage final du modèle transféré *M2* est nécessaire pour l'adapter à une nouvelle base de données *B2*. Il existe trois stratégies différentes de réglage fin pour le modèle *M2*, en fonction de la similarité entre les bases de données *B1* et *B2* ainsi que la taille de *B2*. i) Entraînement complet du réseau (*B1*, *B2* différentes et *B2* importante). ii) Entraînement partiel du réseau : peu de couches (*B1*, *B2* similaires et *B2* importante) ou beaucoup de couches (*B1*, *B2* différentes et *B2* réduite). iii) Entraînement de la classification (*B1*, *B2* similaires et *B2* réduite). En outre, le réseau sera en mesure de fournir de meilleurs résultats si le modèle pré-entraîné présente une diversité de classification et a été entraîné sur une grande base de données. L'apprentissage par transfert a déjà été appliqué pour l'identification de modèle de caméras [1].

2.2 Architectures utilisées

Avec l'émergence de l'apprentissage profond au cours de la dernière décennie, plusieurs défis pour le traitement d'images ont eu lieu, conduisant à l'implémentation de nouvelles architectures de RNC. Certaines sont devenues des standards pour les applications de traitement d'images notamment grâce à leurs performances. Nous avons décidé d'utiliser trois architectures présentant des aspects différents : VGG19 [3] qui est un RNC classique ; ResNet50 [8] qui est un RNC utilisant des sauts de couches pour élargir le domaine des caractéristiques étudiées ; et DenseNet201 [9] qui est un RNC connectant chaque couche avec les suivantes par des sauts de couches, permettant d'obtenir des caractéristiques plus complètes et diverses. Pour chaque architecture, nous avons remplacé la partie classification par une couche d'aplatissement, deux couches denses (de 1028 et 512), deux couches d'abandon (de 0.5) et une sortie de taille N (le nombre de caméras).

2.3 Protocole d'évaluation

TABLE 1 – Description du nombre de modèles de caméras j utilisées pour les modèles pré-entraînés, k inconnues et l totales utilisées pour l'évaluation.

Protocole	**Pré-entrainement**	**Evaluation**	
		Inconnues	**Total**
Expérience 1	j = 8	k = 0	l = 8
Expérience 2	j = 8	k = 8	l = 8
Expérience 3	j = 8	k = 19	l = 27

L'aspect important à considérer pour le protocole d'évaluation est le nombre de modèles de caméras utilisé pour l'apprentissage des caractéristiques et pour la classification. Soit j les caméras utilisées pour l'apprentissage, k et l respectivement les caméras inconnues et le total des caméras pour l'évaluation. Nous avons traité deux sujets réputés de la littérature : l'unicité des empreintes numériques et les caméras inconnues. Tout d'abord, nous avons obtenus des réseaux de référence (un par architecture) à partir d'un ré-entraînement de réseaux pré-entraîné sur ImageNet (détection d'objets). Notre protocole est une série de trois expériences (voir Tab. 1) utilisant les approches de réglage fin. i) Dans un premier temps, nous avons réalisé une évaluation simple de performance grâce à l'apprentissage par transfert avec un entraînement complet du réseau. ii) Puis, nous avons abordé le problème de caméras inconnues. iii) Enfin, la dernière évaluation prend en compte les deux aspects étudiés (caméras inconnues et unicité des empreintes numériques). Le protocole est basé sur l'apprentissage par transfert, dont nous avons présenté les trois possibilités pour régler finement le réseau dans la section 1. Le réglage final du réseau est une approche essentielle de l'apprentissage par transfert et nous avons donc décidé d'évaluer ces trois approches pour déterminer leurs impacts sur les performances des méthodes à l'aide de notre protocole. Les résultats de cette étude sont détaillés dans la section 3.

3 Evaluation expérimentale

3.1 Base de données

Afin d'effectuer une comparaison équitable, tous les réseaux de référence ont été entraînés avec la base de données d'images de Dresde [7]. Elle contient un total de 27 modèles de caméra pour plus de 14 000 images, à partir desquelles nous avons extrait des patchs de taille 128×128 pour les adapter aux entrées des réseaux. Ainsi, le jeu de données final est constitué de 2,6 M de patchs, que nous avons divisés en 3 sous-ensembles : 1,56 M de patchs pour l'entraînement (60%), 0,52 M pour la validation (20%) et 0,52 M pour le test (20%). Pour l'apprentissage, nous

TABLE 2 – Description des sous-ensembles de données pour chaque expérience.

	Transfert	**Validation**	**Evaluation**	**Total**
Exp. 1	448K	149K	149K	746K
Exp. 2	278K	93K	93K	464K
Exp. 3	258K	86K	86K	430K

avons utilisé comme rappel l'arrêt précoce (fin de l'apprentissage si aucune amélioration) et comme optimiseur la descente de gradient stochastique (SGD). Les réseaux ont été entraînés avec une taille de lot de 32 patches et deux GPUs GeForce RTX 2080. Les jeux de données utilisées pour les entraînements ou les transferts de réseaux sont répertoriées dans la Tab. 2.

3.2 Résultats

Dans un premier temps, nous avons réalisé une étude préliminaire dans le but de montrer le problème d'identification lié aux modèles inconnus à l'aide de notre protocole. Notamment nous avons comparé les performances de la première expérience, considérée comme classique (classification de huit modèles connus), avec les deux autres, portées sur les caméras inconnues. L'entraînement des architectures a été effectué avec un apprentissage par transfert et un réglage fin de la partie de classification des réseaux. Les résultats obtenus pour les trois architectures (voir Tab. 3) montrent une perte de performance d'environ 10% (métrique de précision) entre l'expérience 1 et 2, dont le nombre de caméras est similaire (huit modèles). Dès lors, nous en avons conclu que cette baisse de performance était due aux modèles inconnus utilisés dans l'expérience 2. Ce phénomène est encore plus accentué pour l'expérience 3 (19 modèles inconnus) avec une diminution de précision d'environ 17% confirmant le problème soulevé.

TABLE 3 – Résultats de précision des trois expériences du protocole d'évaluation pour une étude préliminaire.

Architectures	*Exp. 1*	*Exp. 2*	*Exp. 3*
VGG 19 [3]	98.47 %	88.52 %	82.66 %
ResNet50 [8]	99.46 %	87.78 %	81.68 %
DenseNet201 [9]	99.49 %	90.02 %	81.82 %
Moyenne	99.14%	88.77%	82.05%

L'étude finale a été mené avec le même protocole pour montrer l'impact des différentes approches sur les performances, mais aussi d'observer le comportement des architectures face au problème de caméras inconnues. Pour l'ensemble des trois expériences, nous avons inclus dans les résultats le temps d'entraînement d'une itération ainsi que la précision des réseaux pour obtenir une comparaison plus complète. Les résultats obtenus montrent que chaque approche de réglage fin des réseaux après l'apprentissage par transfert possède des forces et des faiblesses (voir Tab. 4). Notamment, l'entraînement complet permet d'obtenir de meilleurs résultats, mais nécessite forcément un temps

TABLE 4 – Présentation des résultats pour les architectures choisies et les approches d'apprentissage par transfert en temps d'entraînement par itération, précision (en %) et le nombre de paramètres à entraîner.

Transfert	**Complet**		**Partiel**		**réglage fin**	
	Min.	Acc.	Min.	Acc.	Min.	Acc.
Expérience 2 (8 modèles de caméras inconnues)						
VGG19	21.3	98.02 %	17.7	**97.69 %**	16.6	88.52 %
ResNet50	21.2	97.65 %	17.7	93.84 %	14.3	87.78 %
DenseNet201	**20.5**	**98.85 %**	**12**	96.97 %	**9.5**	**90.82 %**
Moyenne	21	98.11%	15.8	96.17%	13.5	89.04%
Expérience 3 (27 modèles dont 19 caméras inconnues)						
VGG 19	104	**93.48 %**	74.3	**91.22 %**	70.7	**82.66 %**
ResNet50	**88.2**	91.02 %	80.8	87.08 %	68.8	81.68 %
DenseNet201	93	92.58 %	**59.3**	90.05 %	**57.8**	81.82 %
Moyenne	95.1	92.36%	71.5	89.45%	65.8	82.05%

d'entraînement plus long qui s'oppose totalement au réglage fin de la partie de classification, plus rapide, mais aussi moins précis. Le réglage fin partiel, réalisé en incluant le dernier bloc de couches d'extraction de caractéristiques (propre à chaque architecture) offre un compromis entre ces deux approches. L'écart précision-durée pour le transfert complet par rapport au réglage fin partiel est plus avantageux pour l'expérience 2 (2% meilleur pour 6 minutes de plus) que pour l'expérience 3 (3% meilleur pour 25 minutes de plus) encourageant à privilégier le réglage fin partiel pour des bases de données plus importantes. En termes de robustesse aux modèles inconnus, l'architecture VGG19 est plus précise, mais plus longue à entraîner s'opposant à l'architecture DenseNet201. Le constat est similaire aux approches de réglage fin : l'écart précision-durée pour VGG19 par rapport à DenseNet201 est plus avantageux pour l'expérience 2 (0.7% meilleur pour 6 minutes de plus) que pour l'expérience 3 (1.2% meilleur pour 15 minutes de plus). Globalement, ces deux architectures sont adaptées à l'apprentissage par transfert et au réglage fin de réseaux pour l'identification de modèles inconnus.

4 Conclusion et perspectives

Cet article présente une étude sur l'apport de l'apprentissage par transfert pour l'identification de modèle de caméras et notamment pour la classification de modèles inconnus, c'est-à-dire non étudiés lors de l'entraînement du réseau. L'analyse a été effectuée avec notre protocole, de trois expériences basées sur les modèles inconnus, pour différentes architectures de RNC (VGG19, ResNet50 et DenseNet201) afin d'observer l'impact du réglage fin de transfert sur leurs performances. Les résultats montrent dans un premier temps, par le biais d'une étude préliminaire la diminution de performance des réseaux pour la classification de modèles inconnus. L'étude finale atteste de la différence de performance (en temps et en précision) en fonction des approches du réglage fin. Notamment, les résultats montrent que le réglage fin partiel (dernier bloc de l'extracteur de caractéristiques et couches de classification) est à privilégier, au détriment d'un entraînement complet, pour des bases de données conséquentes afin de bénéficier d'un compris efficace. Ce même constat est observable pour les architectures RNC en privilégiant DenseNet201 plutôt que VGG19.

Références

[1] M. H. Al Banna, M. Ali Haider, M. J. Al Nahian, M. M. Islam, K. A. Taher, and M. S. Kaiser. Camera model identification using deep cnn and transfer learning approach. In *2019 International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)*, pages 626–630, 2019.

[2] B. Bayar and M. C. Stamm. Towards open set camera model identification using a deep learning framework. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2007–2011, 2018.

[3] Yoshua Bengio and Yann LeCun, editors. *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[4] Alexandre Berthet and Jean-Luc Dugelay. A review of data preprocessing modules in digital image forensics methods using deep learning. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 281–284, 2020.

[5] L. Bondi, L. Baroffio, D. Guera, P. Bestagini, E. J. Delp, and S. Tubaro. First steps toward camera model identification with convolutional neural networks. *IEEE Signal Processing Letters*, 24(3) :259–263, March 2017.

[6] T. Filler, J. Fridrich, and M. Goljan. Using sensor pattern noise for camera model identification. In *2008 15th IEEE International Conference on Image Processing*, pages 1296–1299, 2008.

[7] Thomas Gloe and Rainer Böhme. The 'dresden image database' for benchmarking digital image forensics. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, SAC '10, pages 1584–1590, New York, NY, USA, 2010. ACM.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[9] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.

[10] C. Li. Source camera identification using enhanced sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 5(2) :280–287, 2010.

[11] Judith A. Redi, Wiem Taktak, and Jean-Luc Dugelay. Digital image forensics : a booklet for beginners. *Multimedia Tools and Applications*, 51(1) :133–162, January 2011.

[12] A. Tuama, F. Comby, and M. Chaumont. Camera model identification with the use of deep convolutional neural networks. In *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, Dec 2016.