Projet Image - Compte-rendu 1

Sécurité Visuelle - Obscuration d'image

Loïc Kerbaul - Valentin Noyé

14 octobre 2024

1 Introduction

L'obscuration des images est une technique essentielle dans divers domaines, dont la protection de la vie privée, la sécurité des données et la manipulation dans un média visuel, avec laquelle nous souhaitons rendre son contenu difficilement perceptible aux personnes qui n'ont pas le besoin ou l'autorisation, que ce soit un objet, un visage ou tout autre donnée confidentielle.

Dans ce compte-rendu, nous présentons notre première étude méthodes d'obscuration de l'image vise à explorer les méthodes plus traditionnelles et modernes d'obscuration, notamment basées sur l'intelligence artificielle et le deep learning. Nous nous baserons sur celui-ci dans les prochains comptes-rendus afin de l'agrémenter ou de décider de l'implémentation.

2 Premier état de l'art des méthodes d'obscuration

Les méthodes d'obscuration peuvent être classées en deux grandes catégories : les méthodes classiques et les méthodes basées sur l'intelligence artificielle.

2.1 Méthodes classiques

Les méthodes classiques se réfèrent aux méthodes du traitement traditionnel de l'image. On y retrouve les techniques telles que le masquage, le floutage ou encore la pixélisation du contenu [2]. L'utilisation de ces techniques sur des images présentent le défaut de ne pas conserver l'apparence initiale de ces dernières, limitant ainsi leurs possibilités d'utilisation. En effet, dans le cadre de la reconnaissance faciale, il est souhaitable de parvenir à conserver suffisamment l'image initiale (par exemple dans le but de la partager), tout en la modifiant de manière à ce qu'une personne mal intentionné ne puisse pas s'en servir afin d'usurper l'identité de quelqu'un. Les méthodes classiques ne parvenant pas à réaliser efficacement ce compromis, il est donc nécessaire de se tourner vers les méthodes utilisant l'intelligence artificielle afin d'avoir de meilleurs résultats (voir plus bas dans ce compte-rendu).

2.1.1 Masquage

La technique de masquage est celle qui dénature le plus l'image originale par rapport aux autres méthodes classiques. Elle consiste à remplacer une zone de l'image par une autre, l'apparence de l'image intiale est donc complétement perdue.

2.1.2 Floutage

La méthode de floutage permet de contrôler l'intensité de l'obscuration de l'image (en jouant sur la force du flou). Un tel effet peut être obtenu par l'application d'un filtre gaussien, ou encore d'un filtre moyenneur.

2.1.3 Pixélisation

Un effet de pixélisation peut aussi permettre l'obscuration d'une image. Cela consiste à diviser l'image en bloc et à remplacer tous les pixels à l'intérieur d'un bloc par une même couleur, décidée selon un critère particulier (par exemple, la moyenne des pixels à l'intérieur du bloc, ou encore en utilisant la couleur la plus fréquente dans le bloc). Tout comme la méthode de floutage, la pixélisation permet de contrôler l'intensité de l'obscuration, en agissant notamment sur la taille des blocs utilisée.

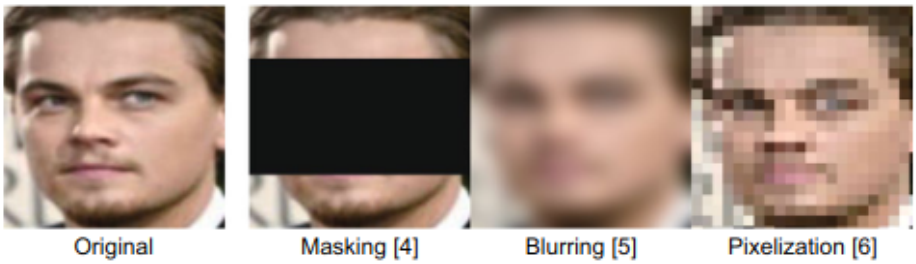

FIGURE 1 – Illustration des différentes techniques classiques d'obscuration

2.1.4 Autres méthodes

Il est possible d'utiliser d'autres méthodes classiques d'obscuration, tel que par chiffrement par clé d'une section de l'image, ou encore par manipulation des pixels ou des bits de cette section, tel que par une permutation, une réecriture des MSB, etc.

2.2 Méthodes par intelligence artificielle

À mesure que les technologies s'améliorent, la robustesse de l'obscuration de l'image devient de plus en plus importante, incitant les chercheurs à réfléchir à des méthodes plus résistantes aux attaques et à la réidentification basées sur l'intelligence artificielle.

2.2.1 Détection du contenu

Il existe plusieurs façons, classiques ou non, de faire de la détection du contenu que l'on souhaite masquer, mais la grande majorité de ces méthodes est basée sur l'intelligence artificielle parce qu'elles ont la forte tendance à détecter de manière très efficace. Par exemple, YOLO [6] constitue l'état-de-l'art de ces techniques.

2.2.2 Manipulation de l'espace latent

Avec un auto-encodeur, l'encodeur compresse les données d'entrée que l'on souhaite masquer dans un espace latent. En manipulant certaines variables latentes (en les mettant à zéro, par exemple), on génère, lors du décodage, des images qui ne ressemblent pas aux originales.

2.2.3 IA générative

Certaines approches à l'offuscation permettent d'offusquer des données confidentielles de manière transparente. NVIDIA (Karras et al.) [3] propose StyleGAN (et plus tard StyleGAN2) permettant de générer des faces à l'aide d'un GAN. D'autres technologies DeepFake ou FaceSwap permettent de remplacer un visage dans une image afin de les anonymiser. Yang et al. [8] discutent d'une méthode de remplacement intelligent d'un texte dans une image.

2.2.4 Suppression du contenu

Parfois, nous voulons simplement supprimer le contenu sensible d'une image. Yu et al. [9] présentent DeepFill, permettant d'offusquer ce contenu par son remplissage intelligent.

3 Premier état de l'art de l'évaluation d'images obscurcies

3.1 Méthodes classiques

3.1.1 Robustesse visuelle

La première méthode est la plus évidente, puisque l'humain est le plus performant en ce qui concerne la reconnaissance du contenu d'une image. L'obscuration d'une image est efficace lorsqu'une personne n'est pas directement capable d'en déterminer son contenu aisément. Plusieurs personnes peuvent être sondées afin de pouvoir nous faire un avis de leur capacité à reconnaître l'objet offusqué.

3.1.2 Métriques de ressemblance

Hanawa et al. [2] utilisent trois métriques de ressemblance par rapport à notre image de base :

— Le PSNR que nous utilisions pour mesurer la ressemblance entre les nuances de deux images (l'originale et la modifiée) ;
— Le SSIM, permettant de mesurer la différence structurelle entre deux images ;
— La métrique LPIPS, permettant de mesurer la ressemblance perceptuelle de deux images.

3.2 Méthodes par intelligence artificielle

3.2.1 Classification par réseau de neurones

Yang et al. [7] définissent plusieurs architectures de CNN utiles à la classification, à savoir AlexNet (Krizhevsky et al.), VGG (Simonyan & Zisserman), SqueezeNet (Iandola et al.), ShuffleNet (Zhang et al.), MobileNet (Howard et al.), ResNet (He et al.), et DenseNet (Huang et al.). Selon les méthodes utilisées, certaines de ces architectures peuvent être capable de déterminer le contenu des images offusquées. AlexNet en l'occurrence est le plus notable de tous car il a gagné le concours de reconnaissance visuelle d'ImageNet.

3.2.2 Réidentification

Orest et al. [4] présentent une méthode de défloutage basée sur des GANs , nommée DeblurGAN. Ces chercheurs [5] présentent DeblurGAN v2, une version plus performante. Dong et al. [1] discutent d'une méthode de dépixelisation par SRCNN (Super-Resolution CNN) en ce qui concerne la dépixelisation d'une section de l'image. Certains modèles de diffusion permettant également de débruiter les sections volontairement altérées d'une image, selon l'intensité de ce bruit. Ces méthodes, combinées à des méthodes classiques d'évaluation de l'obscuration

de l'image (visuelle ou algorithmique) permettent d'en déterminer la robustesse des méthodes utilisées.

Références

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2) :295–307, 2016.

[2] Goki Hanawa, Koichi Ito, and Takafumi Aoki. Face image de-identification based on feature embedding. *EURASIP Journal on Image and Video Processing*, 2024(1) :1–20, 2024.

[3] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *arXiv preprint arXiv :1812.04948*, 2018.

[4] Orest Kupyn, Vitaliy Budzan, Dmitry Mishkin, and Jiri Matas. Deblurgan : Blind motion deblurring using conditional adversarial networks. *arXiv preprint arXiv :1711.07064*, 2018.

[5] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2 : Deblurring (orders-of-magnitude) faster and better. *arXiv preprint arXiv :1908.03826*, 2019.

[6] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once : Unified, real-time object detection. *arXiv preprint arXiv :1506.02640*, 2015.

[7] Kaiyu Yang, Jacqueline Yau, Li Fei-Fei, Jia Deng, and Olga Russakovsky. A study of face obfuscation in imagenet. *arXiv preprint arXiv :2103.06191*, 2021.

[8] Qiangpeng Yang, Hongsheng Jin, Jun Huang, and Wei Lin. Swaptext : Image based texts transfer in scenes. *arXiv preprint arXiv :2003.08152*, 2020.

[9] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5505–5514, 2018.