Université de Montpellier
Master 2 - IMAGINE

Safe-Eye

Évaluation de la sécurité visuelle d'images obscures par CNN

Compte rendu — 2
Projet informatique — HAI927I

Étudiants :
M. Florentin DENIS
M. Khélian LARVET

Encadrants :
M. Nicolas DIBOT
Mme. Bianca JANSEN VAN RENSBURG
M. William PUECH

Année : 2022

Déroulement du projet

Avancement actuel

Pour le moment nous approfondissons nos connaissances sur l'anonymisation et les processus d'obscurcissement utilisés pour des données multimédias. Nous avons ainsi parcouru plusieurs articles et documents afin de réaliser notre propre état de l'art.

Nous commençons également à implémenter des fonctions avec plusieurs paramètres permettant d'obscurcir plus ou moins une image. Voici nos procédés choisis :

- **Pixélisation** : En fonction de la taille du bloc, nous remplaçons le pixel courant et ses voisins par une valeur moyenne.
- **Flouatage** : En fonction de la taille du noyau, nous remplaçons la valeur du pixel courant par une moyenne des pixels voisins.
- **Masquage** : Remplace aléatoirement des pixels dans l'image par un pixel noir, ou un pixel d'une couleur aléatoire.
- **Bruitage par modification de bit** : Permet de modifier tous les pixels de l'image en remplaçant n bits de poids faible, ou fort, par des 0, des 1 ou une valeur binaire aléatoire.

Tâches prévues

Pour la semaine prochaine, nous prévoyons de finir l'implémentation des fonctions d'obscurcissement. Nous essaierons d'analyser les résultats obtenus avec différentes mesures qualitatives en partant des métriques objectives puis en proposant des métriques psycho-visuelles : MSE, PSNR, SSIM, HaarPSI.

Si nous sommes efficaces, nous pourrons également mettre en place un classifieur permettant de reconnaître des visages sur une image.

Difficultés rencontrées et questionnements

Nous ne savons pas encore comment mesurer la robustesse d'une méthode non destructrice face à des algorithmes réversibles. En effet, des attaquants peuvent retrouver les données originales d'un contenu obscurci.

Nous nous posons également des questions concernant la capacité de reconnaissance du CNN. La reconnaissance d'un visage par le CNN est-elle suffisante ou devons nous également faire de la reconnaissance sur des particularités faciales (nez, bouche, yeux, etc.).

État de l'art

L'obscurcissement des images

L'évolution des technologies actuelles, notamment avec les réseaux à haut débit ou encore les appareils électroniques portables abordables (smartphones, appareils photo), ont permis à l'utilisateur moyen de créer et partager des données multimédias avec aisance.
Cette simplification entraîne une augmentation considérable du volume des données multimédias disponible en ligne ainsi qu'une augmentation du risque d'atteinte à la vie privée. De plus, ces données multimédias sont souvent stockées sur des bases de données tierces non fiables, ce qui fait de la confidentialité et de l'anonymat une préoccupation actuelle majeure.

Des techniques d'obscurcissement des images ont ainsi été développées pour protéger les informations sensibles des images. Plus précisément, le principe d'obscurcissement pour une image consiste à modifier voire supprimer des éléments sensibles tout en conservant certaines caractéristiques visuelles permettant son traitement.
Ainsi, ces techniques sont plus ou moins efficaces selon un compromis *intimité-utilité* caractérisé par les éléments spécifiques suivants [3] :

— **La réversibilité** : Cette propriété est obtenue en utilisant des processus de translation pour passer d'un état à un autre. Ces processus doivent rester secrets sinon la sécurité de l'obscurcissement est compromise.
— **L'utilité** : Cette propriété est obtenue en divulguant des informations précises intentionnellement sur le contenu sans dévoiler son entièreté.
— **La naturalité** : Cette propriété est obtenue de manière subjective et permet de dissimuler le fait que l'image ait été obscurci (ce qui peut être un objectif de sécurité en soi).

Ces techniques peuvent être répertoriées dans les trois catégories suivantes :
— **La déformation** : permet de modifier l'élément sensible pour le rendre inintelligible.
— **Le remplacement** : permet de modifier l'élément sensible par un élément visuellement similaire.
— **La suppression** : permet de retirer complètement l'élément sensible.

La suppression assure une intimité absolue mais l'utilité est compromise car aucune information ne pourra en être déduite. À l'inverse, le remplacement et la déformation peuvent divulguer des informations sur le contenu obscurci tout en obtenant une certaine intimité.

Ainsi, plusieurs techniques existent : Protection par chiffrement, "Inpainting", Morphing de visage, Transfert de style, De-identification, etc. Dans notre cas nous utiliserons principalement des techniques de filtrage d'images et de masquage à savoir :

— **La pixélisation** : On remplace le pixel courant et ses voisins par une valeur moyenne.
— **Le floutage** : On remplace le pixel courant par une moyenne des pixels voisins.
— **Le masquage** : On remplace aléatoirement des pixels dans l'image par un pixel noir, ou un pixel d'une couleur aléatoire.
— **Le bruitage** : On remplace le pixel courant par une valeur binaire ayant subi des modifications sur les bits de poids forts ou faibles.

L'évaluation de la sécurité visuelle des images

Une double mesure est nécessaire pour mesurer la sécurité visuelle pour une méthode d'obscurcissement donnée. En effet, certaines méthodes ne sont pas efficaces contre les CNN qui peuvent aisément retrouver et reconnaître (voire reconstruire) une information confidentielle. À l'inverse, certaines méthodes ne sont pas efficaces contre l'oeil humain qui peuvent aisément supposer le contenu initial confidentiel.

Selon le type de contenu sensible que l'on cherche à obscurcir, des évaluations différentes peuvent être proposées [3] :

- **Biométrique** : Unique, mesurable et permanent, permettant une reconnaissance individuelle accrue (Visage, iris, empreinte digitale, etc.).
- **Biométrique léger** : Caractéristiques personnelles qui ne sont pas nécessairement uniques ou permanentes (sexe, couleur de peau, tatouage, etc.).
- **Non biométrique** : Informations contextuelles temporaires et modifiables (plaque d'immatriculation, carte de crédit, etc.).
- **Confidentiel** : Attributs non personnels dissimulés pour des raisons de sécurité ou de confidentialité
- **Censuré** : Censure visuelle du contenu en raison de lois et règlements.

La plupart des images ont plusieurs types contenus sensibles qui doivent être identifiés et protégés. Ainsi, même si le visage d'une personne est obscurci dans une image, des données biométriques légères peuvent être laissées dans l'image. Ces données telles que la silhouette du corps ou la couleur de peau peuvent être des indices importants dans l'identité d'une personne et représente le problème d'identification multimodale.

Pour comparer le degré de confidentialité fourni pour une méthode obscurcissement, nous pouvons utiliser des métriques de similarité entre l'image originale et l'image obscurcie [9] :

- **PSNR (Peak Signal to Noise Ratio)** : Le PSNR est exprimé à l'aide de l'échelle des décibels, et les valeurs typiques pour des images de bonne qualité se situent entre 30 et 50 dB. Par conséquent, les valeurs inférieures au seuil inférieur indiquent que l'image est protégée contre la perception humaine.

- **SSIM (Structural Similarity Index Measure)** : Le SSIM est une métrique de qualité d'image qui peut quantifier la confidentialité des images. Il prend en compte les phénomènes perceptifs tels que la luminosité, le contraste, ainsi que les changements d'informations structurelles. SSIM peut prendre des valeurs comprises entre 0 et 1, où 0 signifie aucune similitude structurelle et 1 indique des images identiques. Par conséquent, les valeurs faibles correspondent à une sécurité accrue.

- **HaarPSI (Haar Wavelet-based Perceptual Similarity Index)** : le HaarPSI est une métrique de similarité pour les images qui vise à évaluer correctement la similarité perceptive entre deux images par rapport à un spectateur humain. Cette mesure obtient des corrélations plus élevées avec les scores d'opinion humain que d'autres mesures de similarité telle que SSIM.

Bibliographie

[1] *Andreea Bianca Popescu, Ioana Antonia Taca, Anamaria Vizitu, Cosmin Ioan Nita :*
(2022) Obfuscation Algorithm for Privacy-Preserving Deep Learning-Based Medical Image Analysis
https://www.researchgate.net

[2] *Christophe Charrier, Chaker Larabi, Hakim Saadane :*
(2005) Evaluation de la qualité des images
https://www.researchgate.net

[3] *Hanaa Abbas, Roberto Di Pietro :*
(2022) Sanitization of Visual Multimedia Content: A Survey of Techniques, Attacks, and Future Directions
https://www.researchgate.net

[4] *Jimmy Tekli, Bechara AL Bouna, Raphaël Couturier, Gilbert Tekli :*
(2019) A Framework for Evaluating Image Obfuscation under Deep Learning-Assisted Privacy Attacks
https://www.researchgate.net

[5] *Alexandre Devaux, Nicolas Paparoditis, Frederic Precioso, Bertrand Cannelle :*
(2009) Face Blurring for Privacy in Street-level Geoviewers Combining Face, Body and Skin Detectors
https://www.researchgate.net

[6] *Slobodan Ribarica, Aladdin Ariyaeeiniab, Nikola Pavesic :*
(2016) De-identification for privacy protection in multimedia content: A survey
https://www.sciencedirect.com

[7] *Elaine M Newton, Latanya Sweeney, Bradley Malin :*
(2005) Preserving privacy by de-identifying face images
https://www.researchgate.net

[8] *Richard McPherson, Reza Shokri, Vitaly Shmatikov :*
(2016) Defeating image obfuscation with deep learning
https://www.researchgate.net

[9] *Rafael Reisenhofer, Sebastian Bosse, Gitta Kutyniok, Thomas Wiegand :*
(2018) A Haar Wavelet-Based Perceptual Similarity Index for Image Quality Assessment
https://www.researchgate.net