Université de Montpellier
Master 2 - IMAGINE

Safe-Eye

Évaluation de la sécurité visuelle d'images obscures par CNN

Compte rendu — 6
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

À l'heure actuelle, notre application possède une interface graphique `Tkinter` et possède de nombreuses fonctionnalités. Celles-ci visent à sécuriser le contenu d'une image mais également à informer l'utilisateur sur la sécurité visuelle des filtres utilisés :

- Importer des images (jpg ou png) pour les modifier et ensuite les exporter (jpg ou png) après obscurcissement.

- Sélectionner des régions d'intérêt manuellement en traçant des carrés directement sur le "canvas" de l'application à l'aide du curseur.

- Appliquer un filtre d'obscurcissement sur un filtre sélectionné, voici les différents filtres possibles :
  - **Pixellisation** avec un paramètre sur la taille des blocs.
  - **Flou Gaussien** avec un paramètre sur la taille du noyau.
  - **Mélange**.
  - **Bruitage aléatoire** avec un paramètre sur la probabilité de remplacer un pixel.
  - **Bruitage sélectif** avec des paramètres sur les bits impactés (LSB/MSB) et sur la valeur de remplacement (0, 1, aléatoire).

- Modifier l'ordre des filtres permettant d'effectuer des combinaisons d'obscurcissement.

- Sélectionner automatiquement des régions d'intérêt via CNN, permettant d'appliquer rapidement des filtres sur plusieurs zones. Nous avons mis en place deux CNN de détection :
  - **MTCNN** (*Multi-task Cascaded Convolutional Networks*) qui est spécialisé sur la reconnaissance des visages humains et des repères faciaux.
  - **YOLOv3** (*You Only Look Once Version 3*) qui est spécialisé sur la reconnaissance d'objets en général.

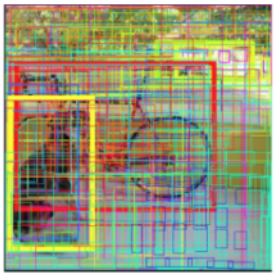
- Attaquer des images obscurcies afin de mesurer la qualité de l'obscurcissement. Nous avons actuellement qu'un seul type d'attaque :
  - **EDSR** (*Enhanced Deep Super-Resolution network*) permettant de créer des images "Super-Résolution".

- Indiquer des mesures de distance visuelle entre l'image originale et l'image modifiée par les filtres dans l'application :
  - **PSNR** (*Peak Signal to Noise Ratio*)
  - **SSIM** (*Structural Similarity Index*)
  - **HaarPSI** (*Haar wavelet-based Perceptual Similarity Index*)

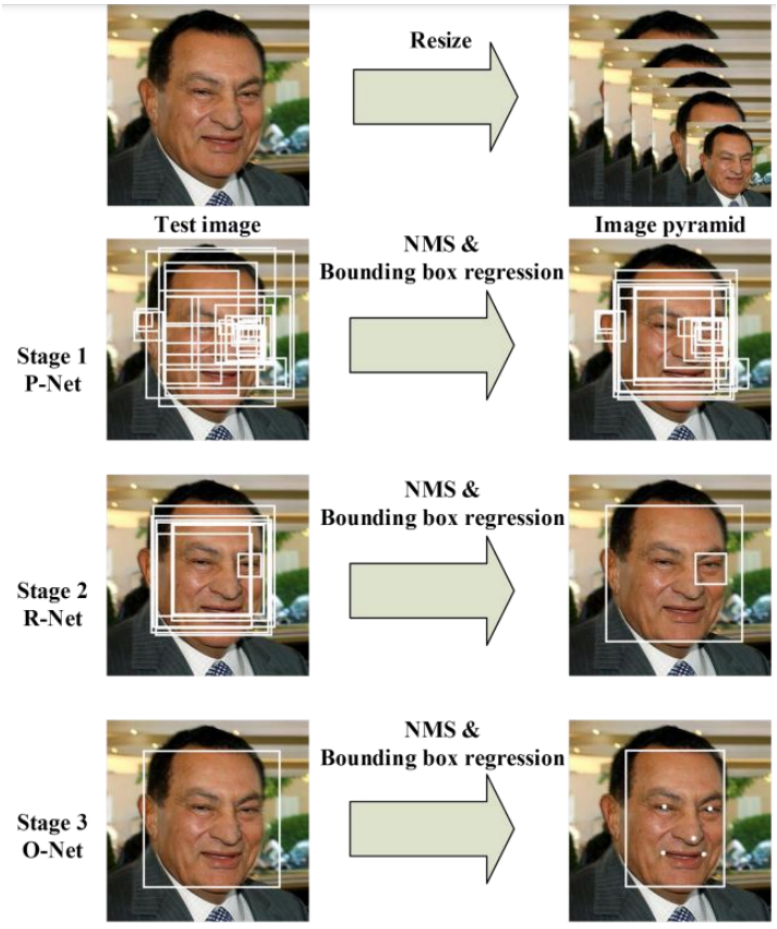
Précisions sur les CNN de détection

YOLOv3 est le premier CNN que nous avons mis en place ayant été entraîné sur la base de données "Coco" (80 classes d'objets). Ce modèle fonctionne en divisant l'image en une grille de cellules où chaque cellule est responsable de la prédiction d'une boîte englobante. Plusieurs boîtes sont ainsi créées pour chaque cellule et les boîtes avec une faible probabilité qu'un objet soit détecté sont supprimées. Enfin, un algorithme "NMS" (*Non-maximum Suppression*) est utilisé pour unifier les boîtes englobantes superposées.






Nous avons par la suite ajouté **MTCNN** pour avoir une détection plus avancée sur les visages humains. Ce modèle fonctionne en utilisant trois autres CNN. Dans un premier temps, P-Net permet de produire rapidement des fenêtres candidates. Ensuite, R-Net affine les fenêtres candidates proposées. Enfin, O-Net permet d'affiner davantage le résultat et d'afficher les positions des repères faciaux.



Précisions sur les attaques

Nous avons pour le moment qu'un seul type d'attaque : l'**EDSR** qui vise à construire une image haute résolution à partir d'une image dégradée. La plupart des études supposent que l'image dégradée est une version sous échantillonnée bicubique mais d'autres facteurs peuvent être pris en compte : flou, décimation, bruit.

Précisions sur les métriques

Le **PSNR** correspond au rapport entre la puissance maximale d'un signal et la puissance du bruit. Cette mesure n'est pas reconnue comme une mesure objective de la qualité visuelle d'une image mais elle nous permet de mesurer un premier indice de distorsion.

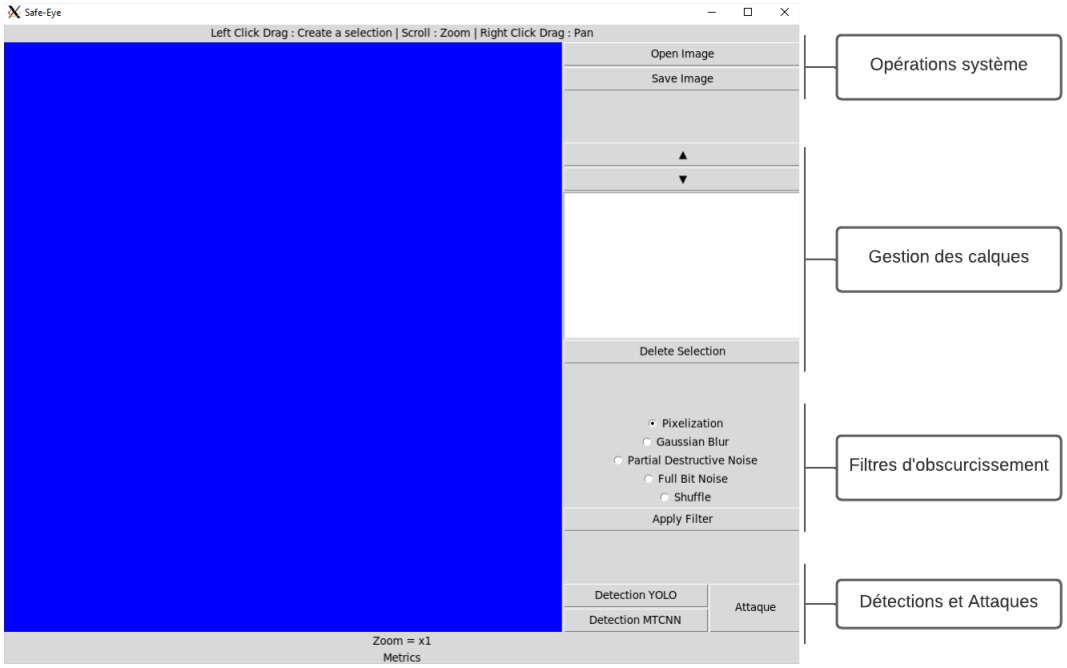
L'idée de **SSIM** est de mesurer la similarité de structure entre les deux images, plutôt qu'une différence pixel à pixel comme le fait par exemple le PSNR. Ainsi, SSIM utilise trois caractéristiques pour comparer des images : la Luminance, le Contraste et la Structure.

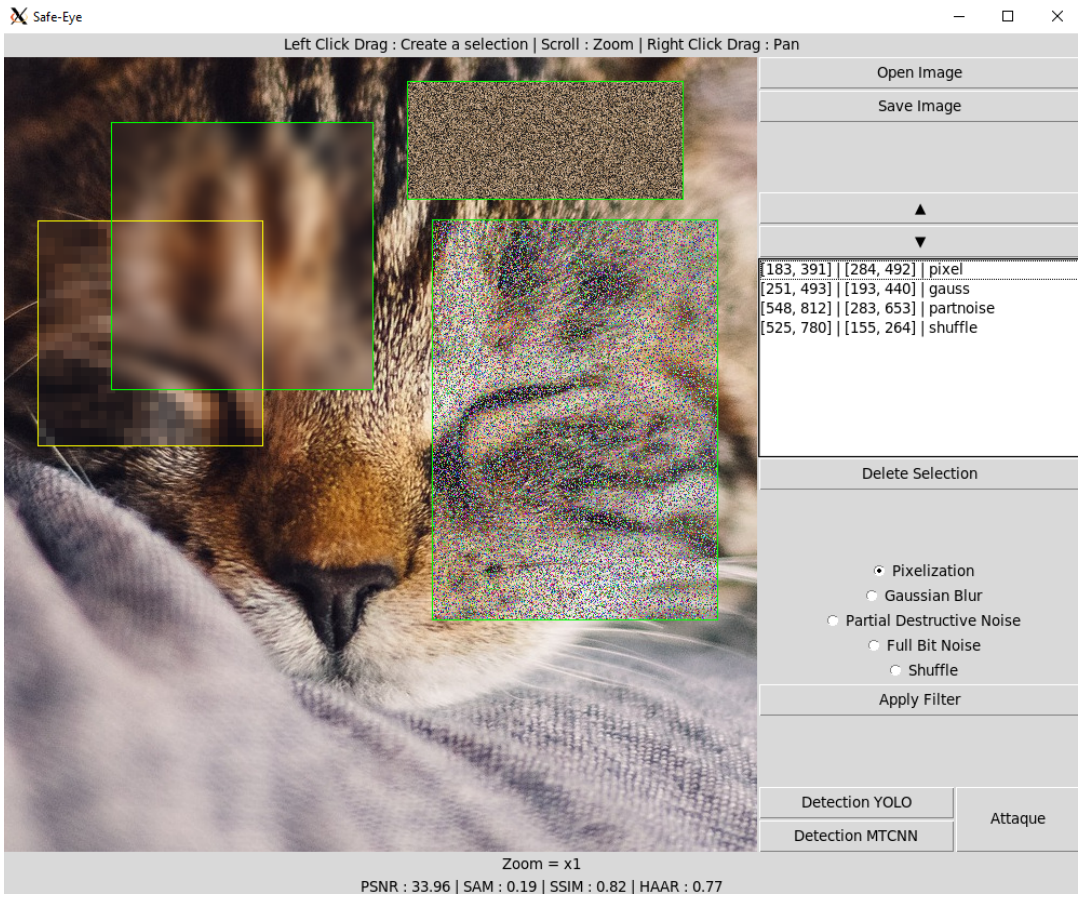
Le **HaarPSI** utilise les amplitudes des coefficients d'ondelettes de Haar à haute fréquence pour définir les similitudes locales et à l'inverse, les coefficients d'ondelettes de Haar à basse fréquence pour pondérer l'importance les dissemblances à des emplacements spécifiques.

Tâches prévues

Nous prévoyons de mettre en place le système d'évaluation utilisateur afin de produire une évaluation de la sécurité visuelle, combinant les mesures quantitatives avec les mesures subjectives des utilisateurs.

Si nous avons le temps, nous essaierons d'ajouter des attaques supplémentaires.


Safe-Eye
Left Click Drag : Create a selection | Scroll : Zoom | Right Click Drag : Pan
Open Image
Save Image
Delete Selection
Pixelization
Gaussian Blur
Partial Destructive Noise
Full Bit Noise
Shuffle
Apply Filter
Detection YOLO
Detection MTCNN
Attaque
Zoom = x1
Metrics
Opérations système
Gestion des calques
Filtres d'obscurcissement
Détections et Attaques


Safe-Eye
Left Click Drag : Create a selection | Scroll : Zoom | Right Click Drag : Pan
Open Image
Save Image
[183, 391] | [284, 492] | pixel
[251, 493] | [193, 440] | gauss
[548, 812] | [283, 653] | partnoise
[525, 780] | [155, 264] | shuffle
Delete Selection
Pixelization
Gaussian Blur
Partial Destructive Noise
Full Bit Noise
Shuffle
Apply Filter
Detection YOLO
Detection MTCNN
Attaque
Zoom = x1
PSNR : 33.96 | SAM : 0.19 | SSIM : 0.82 | HAAR : 0.77

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

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

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

[9] *Slobodan Ribarica, Aladdin Ariyaeeiniab, Nikola Pavesic :*
(2016) De-identification for privacy protection in multimedia content: A survey

[10] *Tanaka, M., Echizen, I., and Kiya, H. :*
(2022) On the Transferability of Adversarial Examples between Encrypted Models

[11] *Hao, H., Guera, D., Horvath, J., Reibman, A. R., and Delp, E. J. :*
(2020) Robustness analysis of face obscuration