Projet Image 2 :
Évaluation de l'obscuration par CNN

Loïc Kerbaul - Valentin Noyé

Master 2 Imagine

Contexte

- Protection de la vie privée
  - Floutage de visage
  - Masquage de données sensibles
- Conformité légale et réglementaire
- Secret-défense
- Protection contre l'ingénierie inverse et l'intelligence artificielle

Comment évaluer la qualité visuelle et la robustesse de nos méthodes d’obscuration?

Plan

1. État de l’art
2. Obscuration
    a. Méthodes traditionnelles
    b. Méthodes par IA
3. Évaluation de l’obscuration
    a. Méthodes traditionnelles
    b. Méthodes par IA
4. Analyse des résultats
5. Démonstration
6. Conclusion

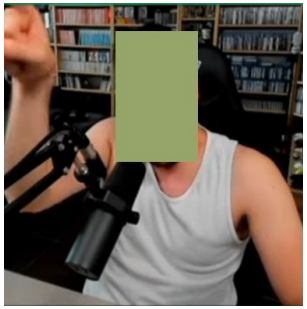
État de l’art

- Méthodes d’obscuration
  - Techniques classiques : Traitement d’images ayant pour but de dénaturer l’image initiale. Processus réversible, par exemple avec des filtres de netteté.
  - Techniques basées sur l’IA : Plus robuste aux attaques et aux tentatives de réidentification, et donc plus efficace que les méthodes classiques. (DeepBlur, DeepPrivacy, ...)

- Évaluation des images obscurcies
  - Métriques de ressemblance : PSNR, SSIM, LPIPS
  - Capacité de réidentification d’un modèle sur une image obscurcie

Méthodes traditionnelles d’obscuration

Masquage

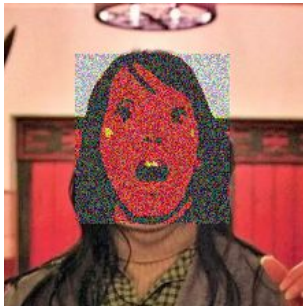


Floutage

Pixélisation

Chiffrement AES

Chiffrement AES partiel



Distorsion spatiale

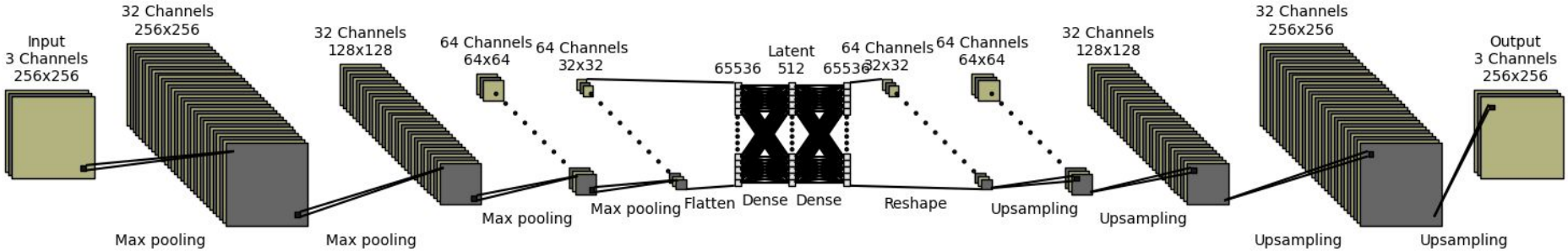
Distorsion colorimétrique

Méthodes d’obscuration par IA

- IA générative
  - Génération de textures ou substitution
  - Transformation d’attributs
- Inpainting et suppression de contenu
- **Approche par auto-encodeur**
  - **Obscuration de l’espace latent**
  - **Modification de la dimensionnalité**

Obscuration par auto-encodeur

- Datasets : **CIFAR-10** puis **LFW**
- Fonction d’activation ReLU, optimiseur ADAM, dropout
- Perte MSE avec un AE, perte KL avec un VAE
- Fitting sur 30 époques avec une taille de batch de 128

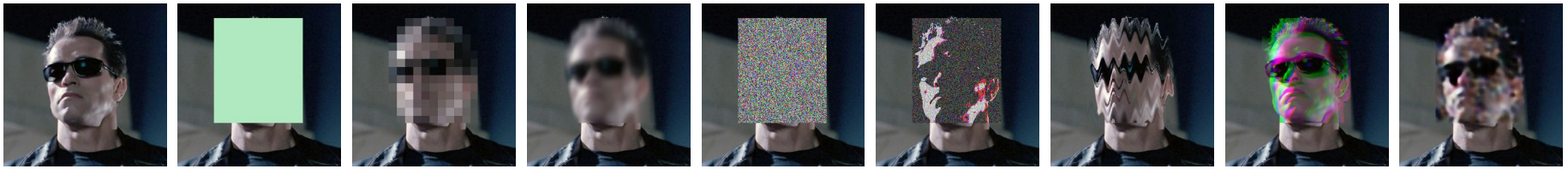


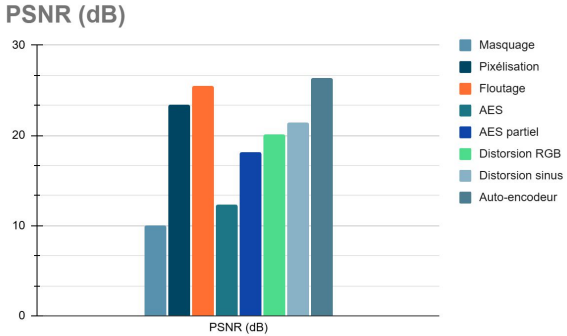
Obscuration par auto-encodeur

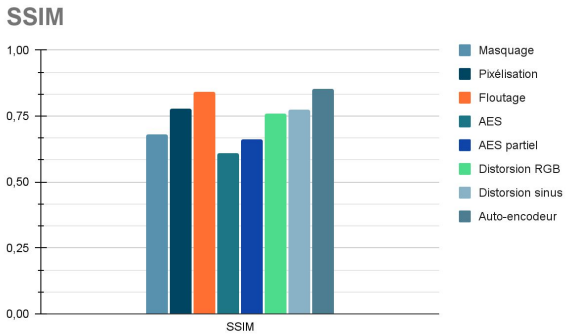
Obscuration de l'espace latent

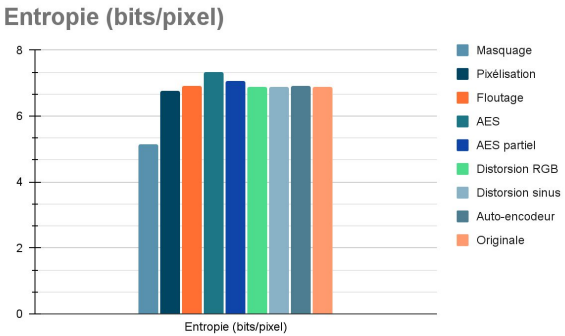
Réduction du nombre de caractéristiques (512)

Méthodes classiques d’évaluation







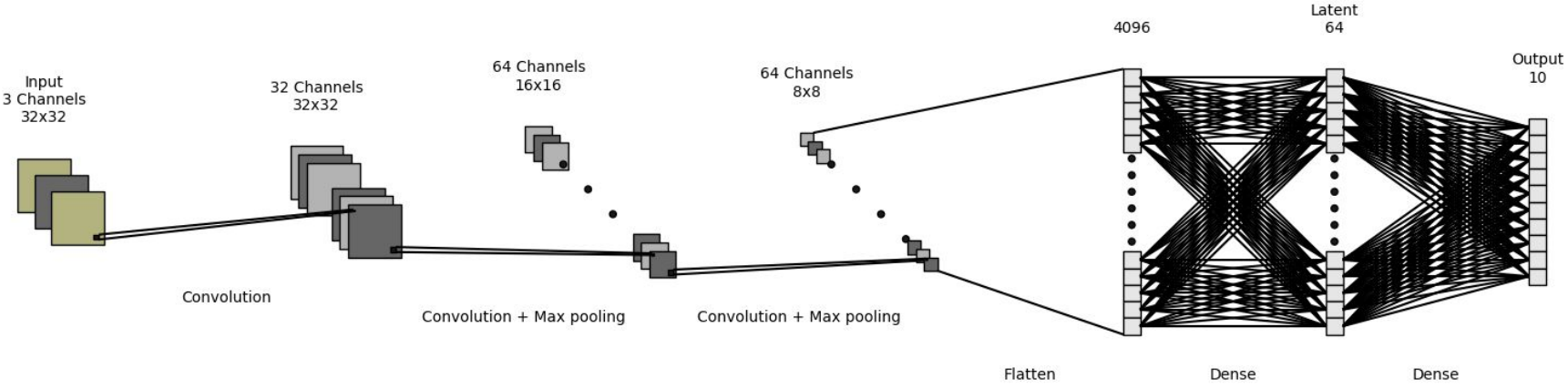


Méthodes d’évaluation par IA

- Fréchet Inception Distance
- Réidentification par défloutage/dépixélisation
- Reconnaissance faciale
- **Classifications par CNN**
  - **Reconnaissance du contenu**
  - **Détection de l’obscuration**
  - **Reconnaissance de la méthode d’obscuration**

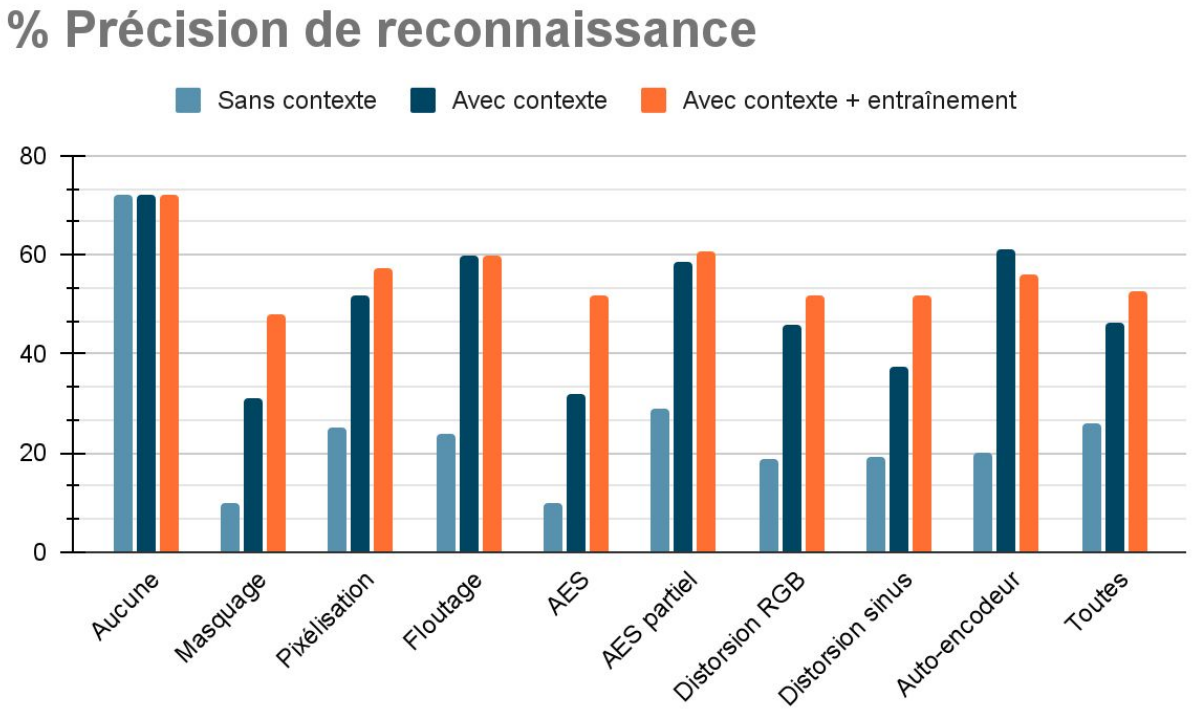
Classification par CNN

- Sur **CIFAR-10** et **Tiny ImageNet**
- Classification contextuelle et non-contextuelle
- Grande plage de paramètres
- ReLU, optimiseur ADAM, perte entropie croisée, sigmoïde ou softmax en fin
- Fitting sur 30 époques, taille de batch de 128 avec arrêt prématuré



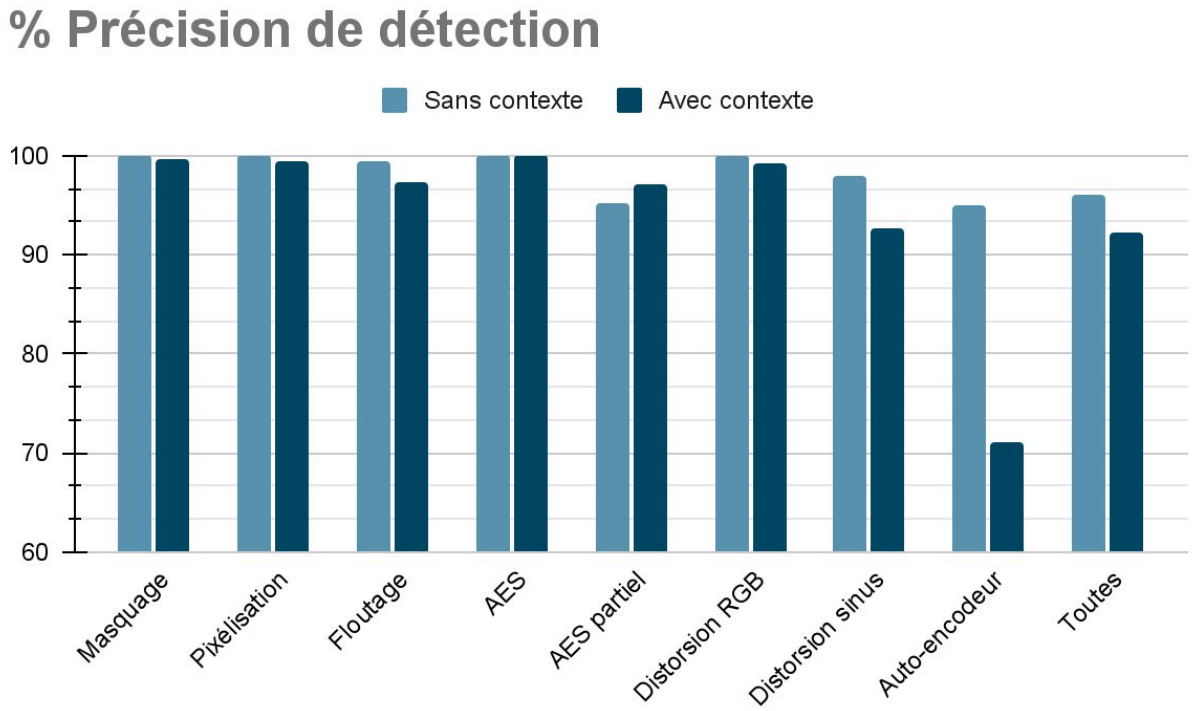
Résultats de la classification

Reconnaissance du contenu



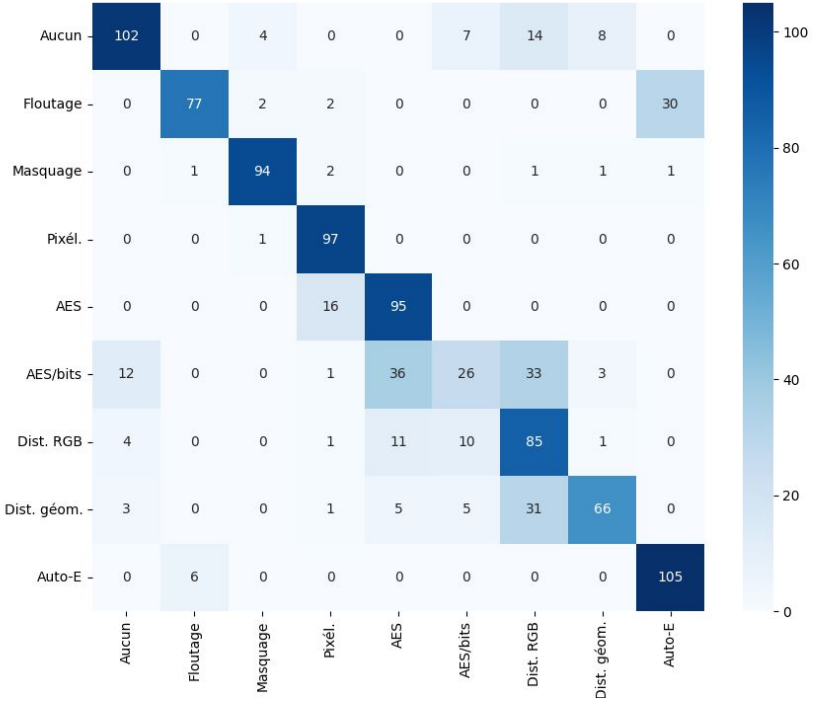
Résultats de la classification

Détection de l'obscuration



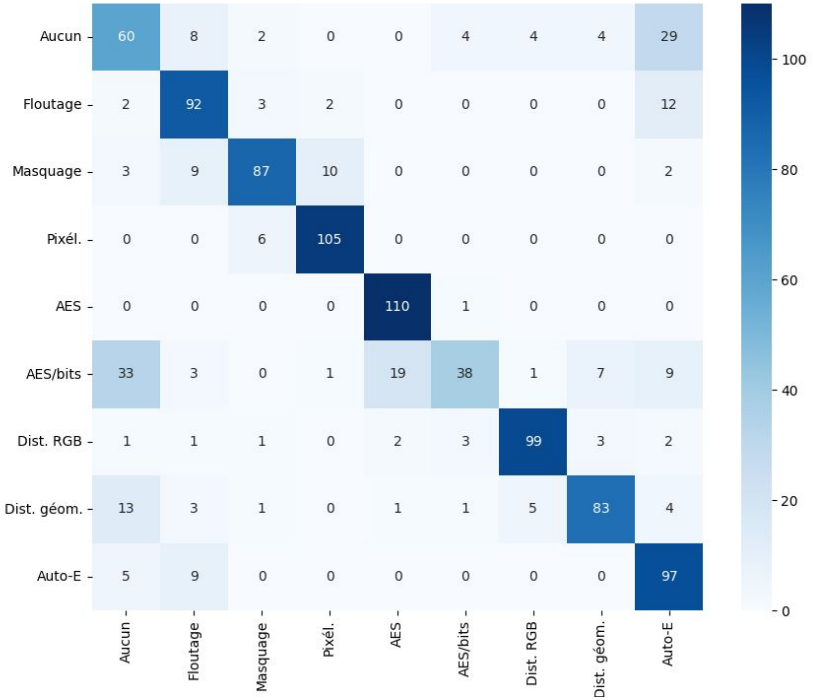
Résultats de la classification

Reconnaissance de la méthode d'obscuration



Sans contexte

(Précision de 74,70%)



Avec contexte

(Précision de 77,10%)

Quelles méthodes d’obscuration choisir?

- Reconnaissance minimale par l’humain et par IA : **Masquage**, **Chiffrement**
- Distorsion minimale : **Auto-encodeur**, **Floutage**, **Pixélisation**
- Inversible sans perte : **Chiffrement**, **Chiffrement de bits**
- Détection minimale par une IA : **Auto-encodeur**

Démonstration

Conclusion et pistes d’amélioration

Références

A Study of Face Obfuscation in ImageNet, Kaiyu Yang, Jacqueline Yau, Li Fei-Fei, Jia Deng, Olga Russakovsky

DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks, Orest Kupyn, Vitaliy Budzan, Dmitry Mishkin, Jiri Matas

Face image de-identification based on feature embedding, Goki Hanawa, Koichi Ito, Takafumi Aoki

DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better, Orest Kupyn, Tetiana Martyniuk, Junru Wu, Zhangyang Wang

Image Super-Resolution Using Deep Convolutional Networks, Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang

A Style-Based Generator Architecture for Generative Adversarial Networks, Tero Karras, Samuli Laine, Timo Aila

Generative image inpainting with contextual attention, Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, Thomas S Huang

SwapText: Image Based Texts Transfer in Scenes, Qiangpeng Yang, Hongsheng Jin, Jun Huang, Wei Lin

Multiscale Spatial-Spectral Convolutional Network with Image-Based Framework for Hyperspectral Imagery Classification, Ximin Cui, Ke Zheng, Lianru Gao, Bing Zhang, Dong Yang, Jinchang Ren

DeepPrivacy2: Towards Realistic Full-Body Anonymization, Håkon Hukkelås, Frank Lindseth

Analysis of Autoencoders for Network Intrusion Detection, Youngrok Song, Sangwon Hyun, Yun-Gyung Cheong

Brain MRI Super-Resolution Using 3D Dilated Convolutional Encoder–Decoder Network, Jinglong Du, Lulu Wang, Yulu Liu, Zexun Zhou, Zhongshi He, Yuanyuan Jia

Generative Adversarial Networks for Synthetic Data Generation: A Comparative Study, Claire Little, Mark James Elliot, Richard Allmendinger, Sahel Shariati Samani