Projet Image M2 : Evaluation de la sécurité visuelle d'images obscures par CNN
HAI918I : Image, sécurité et deep learning

Université de Montpellier - FDS
2ième année Master IMAGINE
Oren AMSALHEM - Thomas CARO

24 Novembre 2024

1 Introduction

Cette semaine nous avons continuer de travailler sur FGSM et avons commencé à travailler sur la vidéo.

2 FGSM

Nous avonsv décidé d'utiliser FGSM sur les images obscurcies de niveau 2 à partir du modèles prenant en compte toutes les méthodes de niveau 2. Voici un résultat de prédiction en lui donnant l'ensemble d'image perturbées par FGSM :

```
Matrice de confusion:
[[   4 3208    5    0 1097]
 [ 456 3663    2    0  194]
 [   2 4290   21    0   14]
 [   0 4292    0    0    0]
 [3181 1049   33    0   40]]

Rapport de classification:
                  precision    recall  f1-score   support

     Image Claire       0.00      0.00      0.00      4314
       Distorsion       0.22      0.85      0.35      4315
Flou de mouvement       0.34      0.00      0.01      4327
    Flou gaussien       0.00      0.00      0.00      4292
     Pixelisation       0.03      0.01      0.01      4303

         accuracy                           0.17     21551
        macro avg       0.12      0.17      0.08     21551
     weighted avg       0.12      0.17      0.08     21551


Précision globale: 17.30%
```

FIGURE 1 – Résultats du CNN avec les images perturbées

En comparaison en envoyant les images obscurcies de base on a :

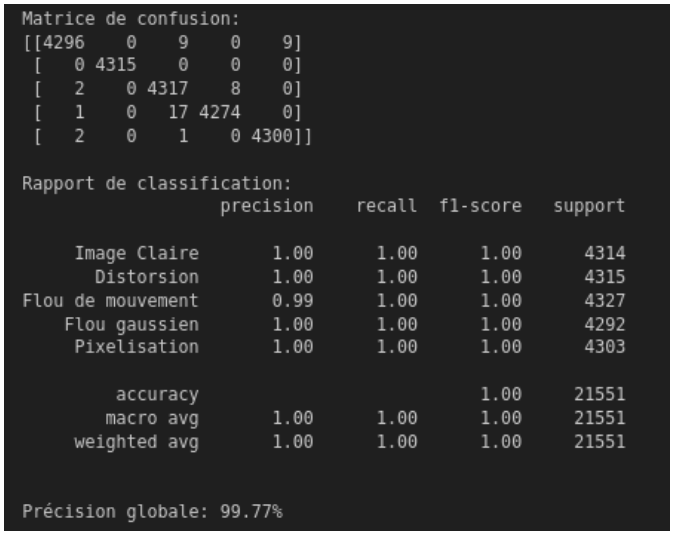

FIGURE 2 – Résultats du CNN avec les images obscurcies de base

Comme on a vu dans le CR précédent nous avons un problème avec les images FGSM de distorsion qui au final arrivent quand même a avoir un f1-score non proche de 0 comme les autres méthodes. Une grande partie des images de distorsion sont bien détectées, par contre une grande quantité des autres types d'images sont aussi catégorisés en distorsion.

Le flou a tendance à uniformiser le gradient de perte de l'image, ainsi les perturbations locales issues de la FGSM ont tendance a donner de grandes difficultés au CNN. Par contre la distorsion garde les caractéristiques de l'image et les déplacent juste selon un certain schéma, sinusoïdal ici. Ainsi les perturbations comme on a pu le voir sur le compte rendu de la semaine dernière sont toujours faites dans les zones de caractéristiques de l'image, ou le gradient est élevé, ainsi la perturbation se fait dans ce sens aussi sur les caractéristiques de l'image correspondant à la distorsion, et les images perturbées ayant comme base la distorsion sont toujours reconnues.

Après plus de tests nous nous sommes rendus compte qu'effectivement les 2 méthodes de flou avec FGSM fonctionnent parfaitement, toutes les images sont mal classées, cependant la méthode de pixelisation avec FGSM ne permet pas de tromper le CNN correspondant car il classe correctement toutes les images. On peut supposer que la raison pour laquelle la pixelisation pour le CNN avec toute méthode avec FGSM fonctionne est que les images sont confondues avec celles de distorsion, mais la méthode en elle même ne performe pas superbement avec FGSM.

3 Vidéo

Nous avons esquissé le plan suivant pour la vidéo du projet :

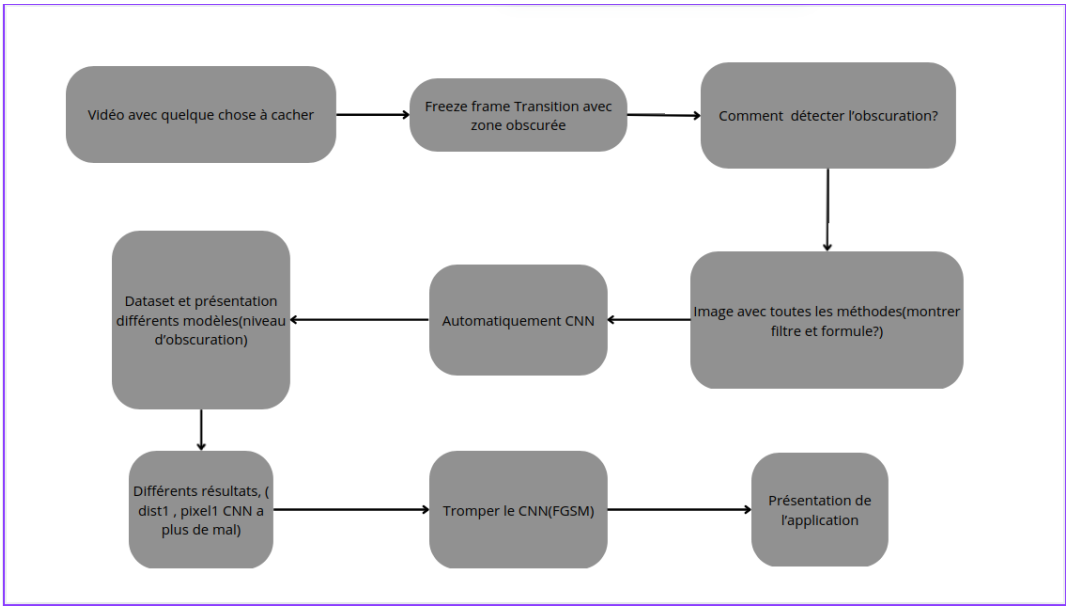

FIGURE 3 – Plan potentiel pour la vidéo

Nous avons commencé à travailler dessus.