



# Compte rendu 6

*BELDJILALI Maxime, CHATEAUNEUF Arthur*

<b>Attaque des informations colorimétriques</b>	<b>2</b>
Niveaux de gris	2
Empoisonnement sélectif des composantes	2
Conclusion sur ces méthodes	3
<b>Reconnaissance par CNN</b>	<b>3</b>
Ajout de flou comme étape de prétraitement :	4
Résultats obtenus sur chacun de nos filtres :	5
<b>Annexes</b>	<b>6</b>
Dépot Github	6

# Attaque des informations colorimétriques

## Niveaux de gris

Nous avons remarqué, dans nos précédents travaux, que les CNN sont sensibles à la teinte des images pour la reconnaissance de classe. Cependant, le système visuel humain, quant à lui, reconnaît un sujet avant tout grâce aux fréquences de luminosité.

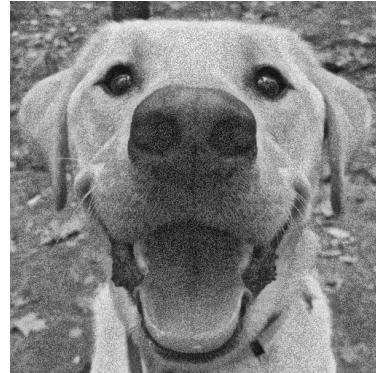
Nous avons ainsi testé l'utilisation d'un filtre en niveau de gris, en le combinant à un bruitage par carte de fréquence :



Niveaux de gris



Bruitage par carte de fréquence  
(opacité max = 40%)



Niveaux de gris avec bruitage  
par carte de fréquence  
(opacité max = 40%)

**Précision CNN : ~47%**

**Précision CNN : ~25%**

**Précision CNN : ~10% (nulle)**

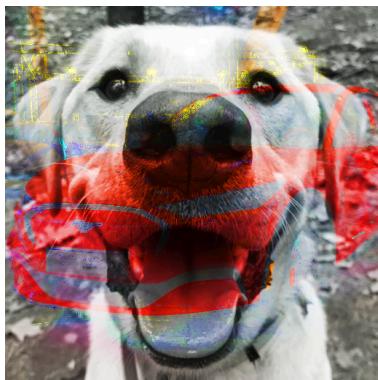
Nous remarquons que la suppression de la teinte et de la saturation permettent de rendre le CNN moins fiable, tout en garantissant une reconnaissance humaine. Nous pouvons ainsi le coupler avec du bruit afin d'affiner les résultats. Nous obtenons ainsi des images perdant de l'information de couleur, mais ayant à subir moins de bruitage afin d'obtenir une précision du CNN égale à l'aléatoire.

## Empoisonnement sélectif des composantes

D'après nos précédents résultats, préserver la cohérence des fréquences de luminosité permet de garder une excellente reconnaissance humaine, tout en laissant aux CNN moins d'outils de reconnaissance. Nous souhaitons, au lieu de simplement effacer les informations de teintes et de saturation, les empoisonner avec des données provenant d'images d'autres classes.

Nous avons ainsi mis au point un filtre qui, pour chaque image donnée, charge une image provenant d'une autre classe. L'image de sortie possède les fréquences de luminosité de l'image originale ainsi que la teinte et saturation de l'image poison.

Nous souhaitons également utiliser notre bruitage par carte de fréquence à différentes opacités afin d'affiner les résultats :



Empoisonnement  
teinte/saturation



Empoisonnement avec  
bruitage par carte de fréquence  
(opacité max = 25%)



Empoisonnement avec  
bruitage par carte de fréquence  
(opacité max = 40%)

**Précision CNN : ~40%**

**Confusion avec le poison :**  
~16%

**Précision CNN : ~10% (nulle)**

**Confusion avec le poison :**  
~18%

**Précision CNN : ~5%**

**Confusion avec le poison :**  
~15%

Nous obtenons des résultats intéressants. L'empoisonnement ne confond pas le CNN pour qu'il reconnaisse le type de classe du poison. Cependant, l'efficacité est très grande. La précision du CNN devient égale à celle de l'aléatoire lorsque l'opacité du bruitage est à 25%.

Cependant, dès lors que l'on applique 40% ou plus d'opacité de bruit, la précision du CNN chute en dessous de celle de l'aléatoire. Dans notre dernier résultat, où la précision du CNN est de 5%, nous remarquons que la classe prédictive est 96% du temps "frog", "cat" ou "bird". Certaines classes, comme "airplane", "automobile" ou "dog" ne sont jamais utilisées dans les prédictions pour ce filtre. De plus, dès lors que la classe réelle de l'image correspond à l'une des 3 classes que le CNN utilise 96% du temps, il ne l'utilise presque plus et obtient les pires scores de toutes classes confondues. Sa précision sur les classes "frog", "cat" et "bird" est ainsi de 0 à 2%.

## Conclusion sur ces méthodes

Nous pensons que l'espace couleur teinte, saturation et luminosité est un très bon candidat pour obtenir des images sécurisées pour une reconnaissance sur les CNN simples. Nous souhaitons continuer d'expérimenter avec ces filtres, afin de mêler bruitage, empoisonnement et manipulation des informations peu sensibles au système visuel humain.

Il faudra également suivre l'évolution de leurs capacités sur les améliorations que nous allons donner à notre CNN.



Plus d'exemples de notre dernier filtre

# Reconnaissance par CNN

## Ajout de flou comme étape de prétraitement :

Dans nos travaux précédents, nous avions entraîné un CNN sur des images 32 par 32 avec le jeu de données Cifar-10. Puis nous avons choisi un ensemble d'images de test supplémentaire de plus haute résolution correspondant plus à une utilisation réelle. Ces images-là font 512 par 512, donc nous devons rajouter une étape de prétraitement pour transformer ces images en images de 32 par 32. Pour cela, nous utilisons une interpolation au plus proche voisin pour réduire la résolution. Puis nous faisons l'inférence sur cette image réduite.

Concernant l'interpolation utilisée, nous avons aussi essayé une interpolation bilinéaire et bicubique mais cela n'avait pas d'impact significatif sur nos résultats à première vue. Donc, nous avons choisi de rester sur une interpolation au plus proche voisin.

D'après les résultats de la semaine passée, nous avons pu constater que les images fortement floutées étaient mieux classées que les non-floutées. Cela semble s'expliquer par une meilleure extraction des basses fréquences, facilitant le travail pour notre CNN. Ainsi, nous avons eu l'idée de rajouter une étape de floutage de l'image dans le prétraitement de nos images. Nous utilisons le même floutage que décrit les semaines précédentes. Nous avons donc une première étape de floutage de l'image, puis une étape de réduction de la résolution avant de faire l'inférence par notre CNN.



Image originale 512 x 512 pixels

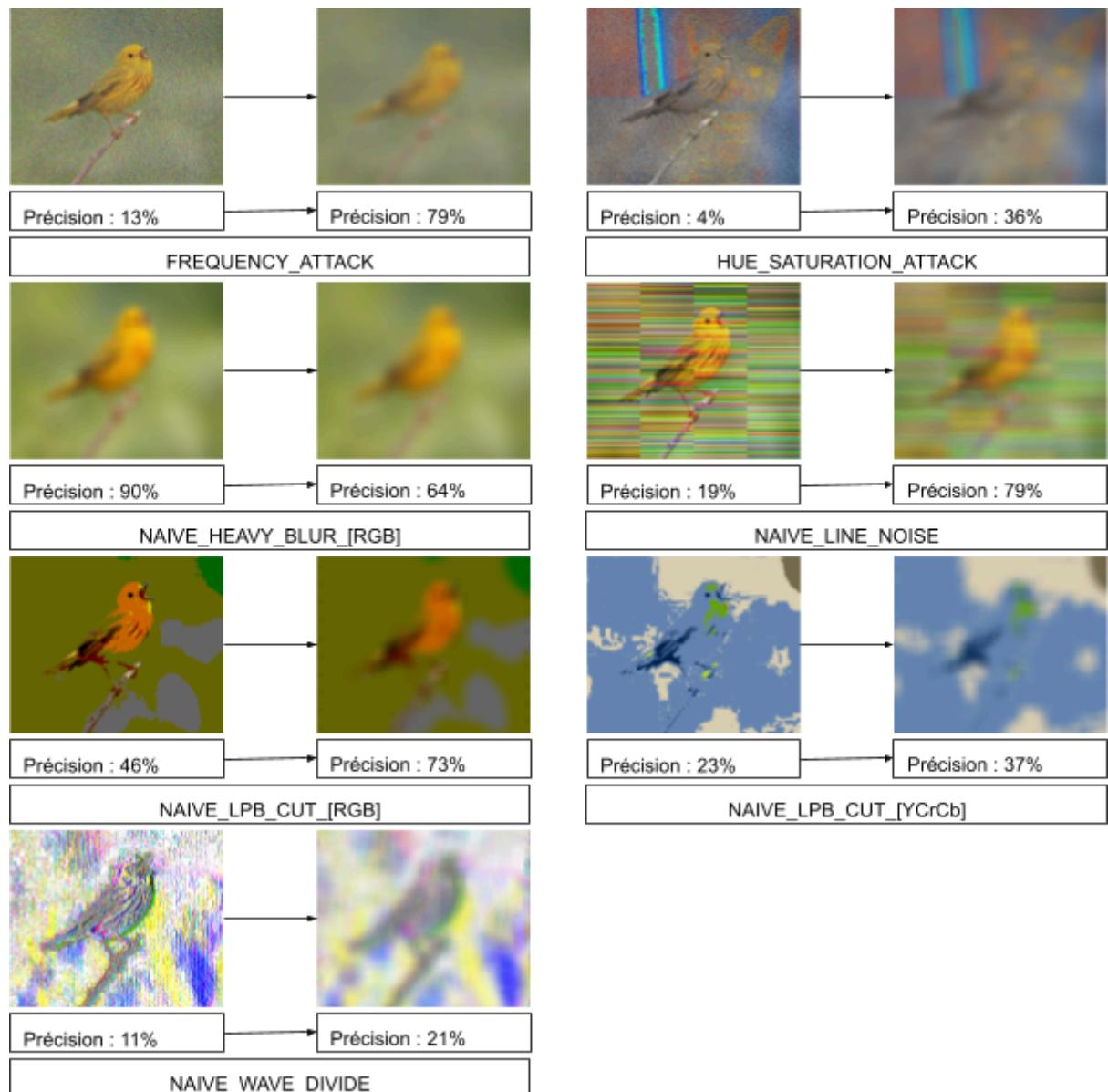
Image floutée de 512 x 512 pixels

Image échantillonnée à 32 x 32 pixels

Nous supposons que cela permettra d'améliorer les résultats dans certains cas en réduisant l'impact des basses fréquences sur l'inférence de notre CNN.

## Résultats obtenus sur chacun de nos filtres :

Nous allons ici faire un comparatif de la précision obtenue pour chacun de nos filtres en rajoutant uniquement cette étape de floutage comme prétraitement. L'étape d'échantillonnage est quant à elle obligatoire et présente dans les deux cas. Ces résultats sont obtenus sur un ensemble de 100 images, avec 10 images pour chacune des classes sur lesquelles nous avons fait notre entraînement. Dans un premier temps, sans floutage en prétraitement et dans un second avec un floutage en prétraitement.



Si on exclut le cas de floutage lui-même, nous constatons qu'il y a une amélioration générale de la précision. Nous obtenons de meilleurs résultats dans les cas où l'action principale du filtre est un bruitage de l'image. À l'inverse, si l'image est déjà floutée, cela n'améliore pas les résultats et au contraire cela dégrade la précision. Ainsi, cela augmente fortement la précision sur les filtres s'attaquant à des fréquences ou des domaines précis de l'image comme la teinte. De plus, nous constatons aussi une amélioration de la précision moindre sur les filtres de coupe de plans binaires et ceux modifiant la couleur.

# **Annexes**

## Dépot Github

- <https://github.com/Pataeon/Securite-visuelle>