



Rapport Visualisation des réseaux de neurones

RDFIA - TP 8

Etudiant :

DJEGHRI Amine

MAMOU Idles

Numéro :

3801757

3803676

Décembre 2020

Partie 1 – Carte de saillance

1- Une ligne des résultats obtenus :

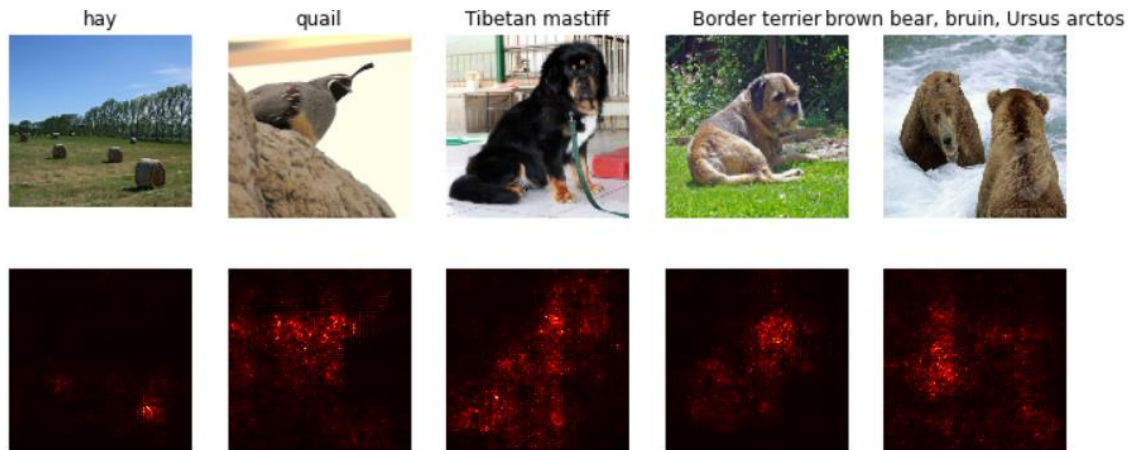


Figure 1 – Images et leurs cartes de saillance

La carte de saillance indique les zones plus importantes d'une image pour prédire la bonne classe. Pour l'image des foin par exemple, on remarque que les pixels les plus importants se trouvent sur les bottes de foin. Pour l'image du chien à droite, c'est la tête du chien.

2- Les limites de cette technique :

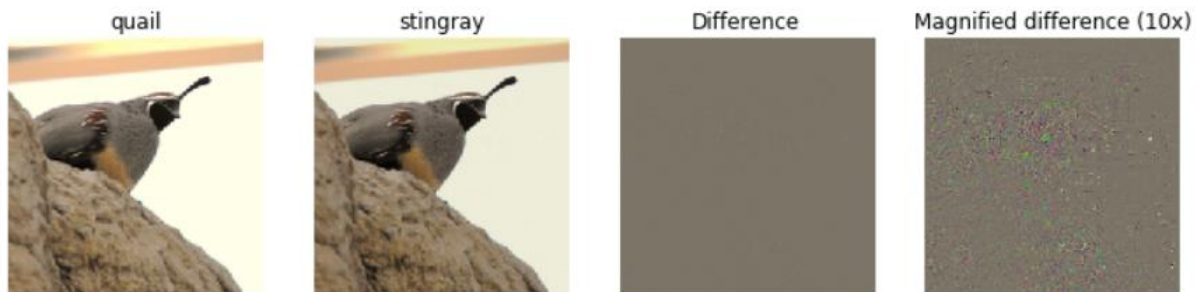
- Ne détermine pas toutes les zones qui sont importantes
- On peut faire en sorte d'avoir une mauvaise prédiction en modifiant les pixels importants non visibles à l'œil nu.
- Indique l'importance de plusieurs pixels au lieu de chaque pixel pour l'identification

3- Oui, la segmentation d'images, localisation d'objets

4- Bonus

Partie 2 – Exemples adversaires

5-



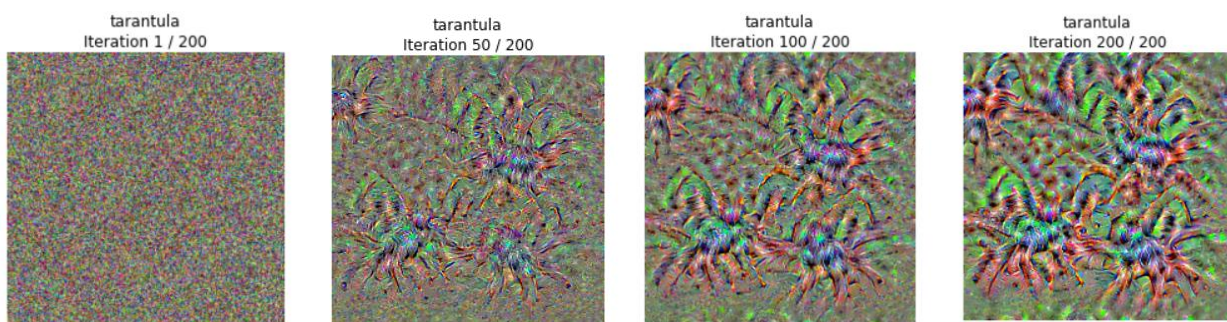
On peut faire en sorte qu'une nouvelle image soit désormais mal classée par le réseau sans que la différence entre les deux images ne soit visible à l'œil nu. Par exemple dans la photo ci-dessus, on est parvenu à faire croire au réseau que l'image d'oiseau est un rayon de la mer, en réalisant une remontée de gradient. La différence peut être clairement grâce à l'image de différence x 10.

- 6- En pratique, l'utilisation de ces réseaux de neurones n'est pas si fiable surtout si le modèle est accessible, car on peut tromper un réseau de neurone et lui faire croire que c'est une autre classe même si l'image est bien classée par un œil humain.

Par exemple : si on modifie une image d'un panneau de signalisation qui n'est pas perceptible par l'œil nu, on pourra tromper les voitures autonomes lors de la détection.

Partie 3 – Visualisation de classes

8-

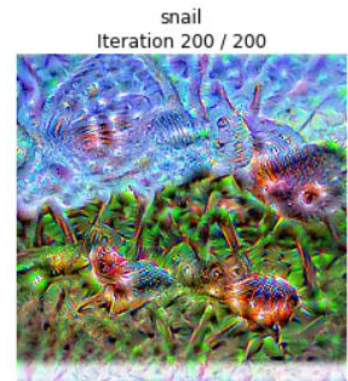
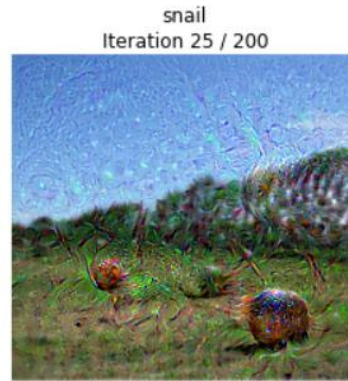


Dans cette partie on cherche à visualiser le type de patterns permettant de produire une classe particulière en sortie du réseau, on veut modifier une image de bruit de façon à maximiser le score de la classe tarantule. Les pattes ainsi que

le corps de la tarentule sont visibles, Plus le nombre d'itérations augmente et plus ces caractéristiques apparaissent fortement dans l'image.

9- Sur le notebook

10-



On utilise maintenant une image d'ImageNet comme image source au lieu d'une image aléatoire, ceci nous permettra de converger plus vite.

On remarque bien la présence des pattes et du corps dans l'image finale, le corps de la tarentule est utilisé sur les bottes de foin de l'image initiale car cette approche utilise les caractéristiques déjà présentes sur l'image.

11- remplacer et tester le modèle avec vgg16