

## Compte Rendu

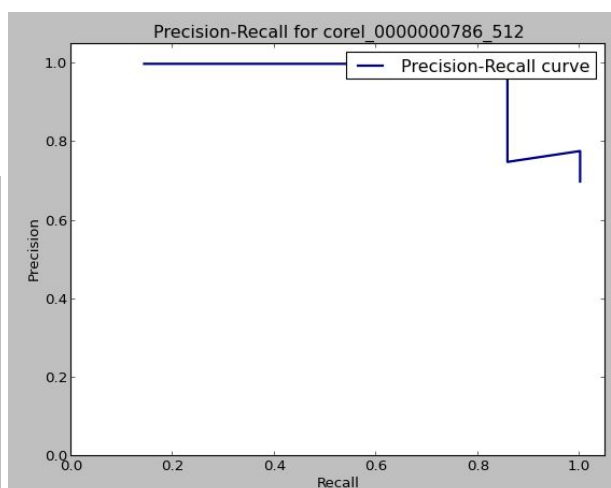
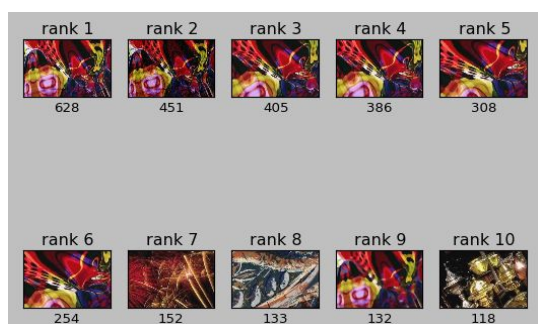
L'objectif du TP est de mettre en place un système de recherche d'images par image. Pour cela nous disposons d'une base de données d'images références et de tests.

Chacune de ces images est décrite par plusieurs vecteurs de caractéristiques obtenu par l'algorithme de SIFT. La recherche d'image revient donc à trouver les similitudes entre ces vecteurs de caractéristiques des images de références et des images de tests. Notre base de données d'images devient donc une base de données de descripteurs.

Pour une image donnée, nous utilisons un knn pour chaque descripteur de l'image et nous en déduisons les plus proches descripteurs de ces derniers. Néanmoins nous n'avons pas mis en correspondance les descripteurs sorti avec leur images, nous avons donc choisi d'établir au préalable une map associant descripteurs et le path de l'image. Une fois cela établi, il nous reste à comptabiliser les descripteurs trouvés correspondant aux images. Pour obtenir les images ressemblantes à celle d'entrée nous effectuons un vote par majorité.

Au lieu de comparer le SIFT nous pouvons utiliser un réseau de neurones (VGG16) afin d'obtenir un descripteur de l'image. Cependant le réseau de neurone fourni un et un seul vecteur différent de ceux obtenu par SIFT pour chaque image. Nous devons donc calculer une nouvelle base de données remplie uniquement de descripteurs issus de VGG16. A noter que nous devons au préalable retirer la dernière couche du réseau de neurone afin de récupérer ce descripteur car la dernière est utilisée pour la classification d'images.

Nous allons maintenant comparer les 2 méthodes en terme de temps et de qualité. Dans un premier temps, pour une image donnée nous obtenons les images suivantes avec SIFT. Ces images sont triées par ordre décroissant de similitude. La première ayant un score de 628, ce qui signifie que 628 descripteurs de cette image sont proches des descripteurs de l'image d'entrée.



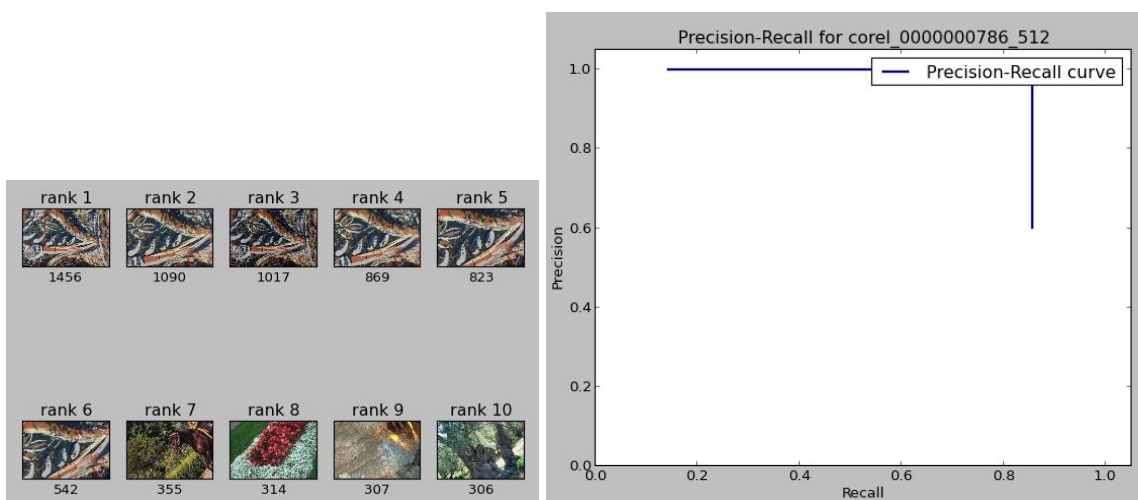
Classement de la recherche pour l'image 1 et Précision-Rappel associé (SIFT)

$Précision(t) = \text{Nombre d'images pertinentes}(t) / \text{Nombre d'images}(t)$

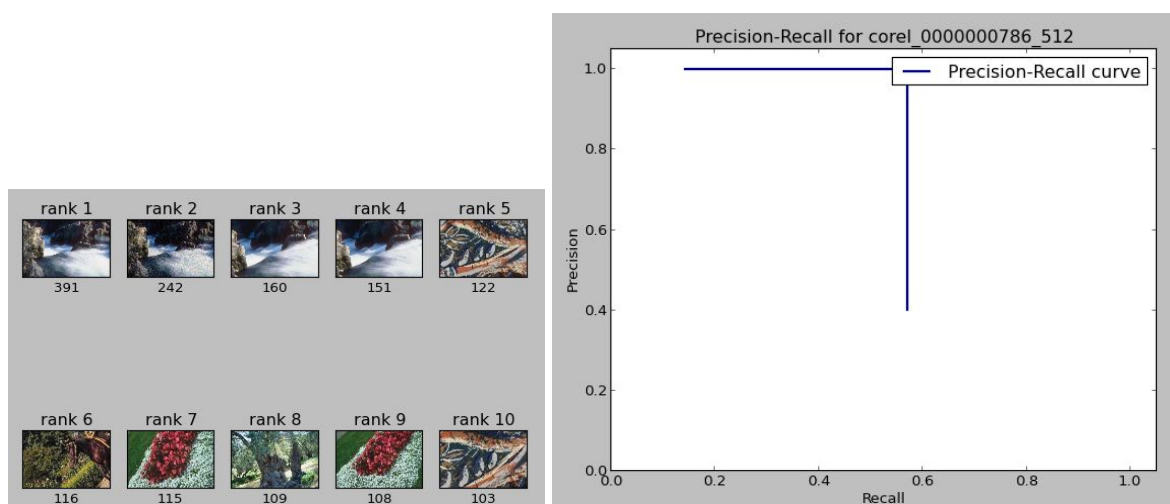
$Rappel(t) = \text{Nombre d'images pertinentes}(t) / \text{Nombre total d'images pertinentes}$

avec  $t$  le nombre d'itérations.

Chaque image testée dispose de 7 images similaires, par conséquent celles-ci devraient figurer dans le top 10. C'est le cas de notre premier résultat même si l'une d'entre-elles est seulement classée à la 9ème position. Le graphique représente la courbe précision/rappel. Nous pouvons voir sur ce graphique que la précision est parfaite (égale à 1) avant de diminuer. En effet, la baisse de la précision est dû à une image pertinente non-trouvée en 7ème position. Elle augmentera légèrement par la suite lorsque la 7ème image pertinente sera trouvée. Néanmoins il arrive qu'elles ne figurent pas toutes, comme c'est le cas pour les 2 résultats suivant.

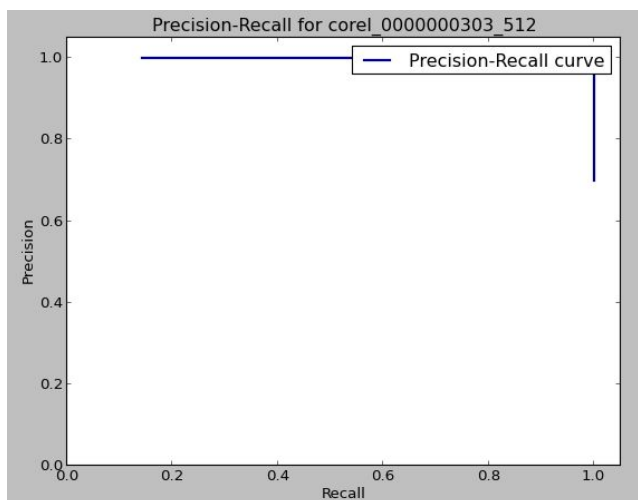


Classement de la recherche pour l'image 2 et Précision-Rappel associé (SIFT)

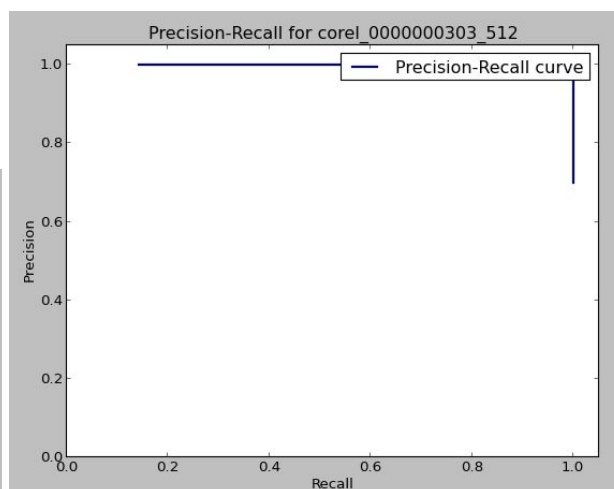
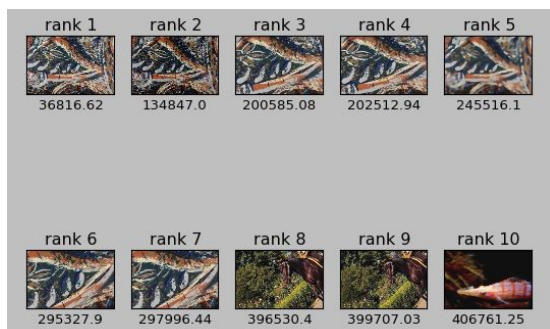


Classement de la recherche pour l'image 3 et Précision-Rappel associé (SIFT)

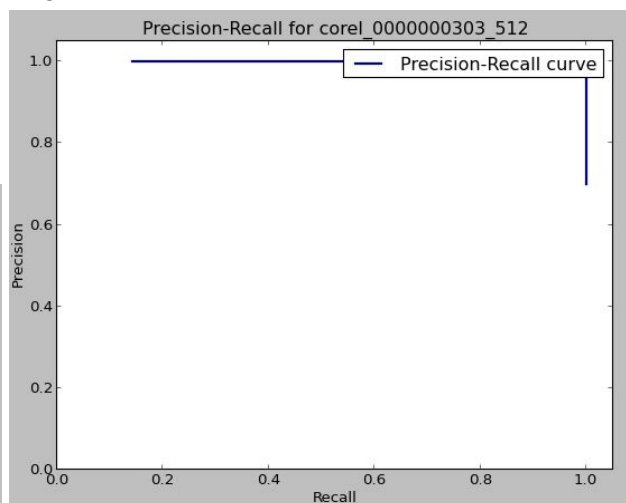
On répète le même processus mais cette fois-ci avec VGG16, les résultats sont bien meilleurs puisque toutes les images se situent dans le top 7. De plus le temps de calcul est bien plus faible que son homologue (se référer au tableau comparatif en bas de page). La courbe précision/rappel, quant à elle, est parfaite. A noter, cette fois ci le score indiqué en dessous de chaque image correspond à la distance entre le vecteur description de l'image d'entrée et celui de l'image résultante.



Classement de la recherche pour l'image 1 et Précision-Rappel associé (CNN)



Classement de la recherche pour l'image 2 et Précision-Rappel associé (CNN)



Classement de la recherche pour l'image 3 et Précision-Rappel associé (CNN)

### Conclusion :

Comme nous pouvons le voir dans ce tableau comparatif, le CNN démontre de bien meilleurs performances que le SIFT en terme de recherche d'images.

	Image 1	Image 2	Image 3
SIFT	7/7*	6/7	4/7
CNN	<b>7/7</b>	<b>7/7</b>	<b>7/7</b>
Temps SIFT (s)	10.95	26.02	8.36
Temps CNN (s)	<b>0.289</b>	<b>0.288</b>	<b>0.321</b>

\* (Les 7 images ne sont pas forcément dans les 7 premières)