

TP : Moteur de recherche d'images

I - Les systèmes d'évaluations

Pour réaliser, un moteur de recherche sur les images nous avons utilisé différentes méthodes :

- Recherches à base d'histogrammes de couleurs: Un histogramme est une courbe statistique indiquant la répartition des pixels selon leur valeur. Nous convertissons les images en histogrammes de couleurs à partir des 3 histogrammes de chacune des composantes Rouge, Vert, Bleu que nous joignons en un. L'objectif grâce à cette méthode est de comparer les valeurs de l'histogramme de la requête et les histogrammes dans la base de recherche pour retrouver si les images décrivent une même scène. L'un des potentiels défaut de cette méthodologie est liée à l'exposition de l'image. En effet, les valeurs de l'histogramme fluctue en fonction de l'éclairage et donc de l'intensité. Il est donc possible de prendre exactement la même photo avec des puissances lumineuses différentes et d'avoir deux histogrammes différents.
- Recherche par sacs de mots visuels: L'utilisation de sacs de mots visuels se fait à partir des descripteurs SIFT (Scale Invariant Feature Transform). Il s'agit d'informations numériques dérivées de l'analyse locale d'une image et qui caractérisent le contenu visuel de cette image de la façon la plus indépendante possible de l'échelle (« zoom » et résolution du capteur), du cadrage, de l'angle d'observation et de l'exposition (luminosité). A partir d'un vocabulaire prédéfini de taille fixe nous pouvons trouver le mot visuel le plus proche pour chaque descripteur. Ainsi, pour chaque image, on a une liste des mots et leur fréquence dans l'image. Pour réaliser une recherche, il faudra comparer la fréquence des mots entre la requête et chaque image de la base de données.

Pour utiliser cette méthodologie, nous avons utilisé plusieurs indexations possibles sur la base de données. Une indexation linéaire, qui est l'utilisation classique d'un index, qui à partir de l'ID d'une image on obtient son sac de mots visuels. Un index inversé, qui lui à partir des mots du vocabulaire, retourne les ID qui contiennent mot visuel.

- Recherche par descripteur VLAD: Les descripteurs VLAD (Vector of Locally Aggregated Descriptors) reprennent le principe des sacs de mots mais donnent plus d'informations sur les descripteurs. Chaque photo est représentée par la somme des résidus entre le mot visuel et son descripteur pour chaque mot du vocabulaire. On obtient donc une matrice de taille vocabulaire x taille descripteur. Il est possible de réaliser une ACP sur chaque matrice afin d'optimiser l'espace mémoire et pour obtenir de meilleurs résultats.

II - Le protocole d'évaluation

Notre objectif, lors de ce TP est d'étudier l'influence des différents paramètres de nos 3 systèmes (norme utilisée, nombre de bins, taille du vocabulaire...) sur la qualité des résultats. Nous baserons notre analyse selon trois principales métriques: le score mAP, les courbes rappel/précision, et le temps moyen de calcul d'une requête.

-Notre évaluation pour l'histogramme est constituée de la variation de la norme, du nombre de bins et du temps de calcul pour une requête.

-Pour le système par sac de mots nous étudions dans un premier temps l'influence du prétraitement de l'index (normalisation/pondération) puis l'influence de la norme choisie sur les résultats. Dans un second temps, nous allons analyser l'influence de la taille du vocabulaire sur la précision du système, ainsi que sur le temps de calcul d'une requête. Enfin, nous ajouterons à notre système un index inversé et étudierons l'influence de celui-ci sur les résultats.

-L'évaluation sur le descripteur VLAD sera basée sur l'influence du nombre de composantes principales obtenus par ACP, puis finira par l'influence de la taille du vocabulaire.

NB: Par manque de performances et de temps, nous avons limité la taille de notre base de données à 500 images ainsi que la taille de nos vocabulaires à 5000 mots pour le système à sac de mots et à 100 mots pour le système VLAD. Nous sommes conscient qu'il serait pertinent d'augmenter le nombre d'images ou la taille du vocabulaire (notamment pour le système à sac de mots) pour pouvoir mieux étudier la convergence des systèmes.

III - Analyse des résultats d'expérimentation

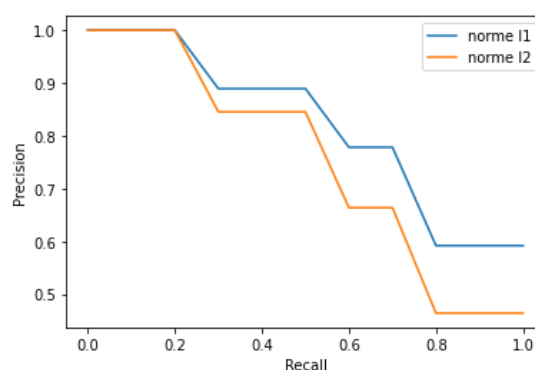
1) Histogramme

- a) Influence de la norme : Nous avons fait varier la norme de calcul des distances entre la norme L1 (Minkowski) et la norme L2 (Euclidienne). Pour ce test, nous avons fixé la taille de chaque "bin" à 12. On observe sur le score mAP et les courbes Rappel-Précision de la norme L1 sont plus performantes. Pour cette raison par la suite nous garderons L1 comme norme.

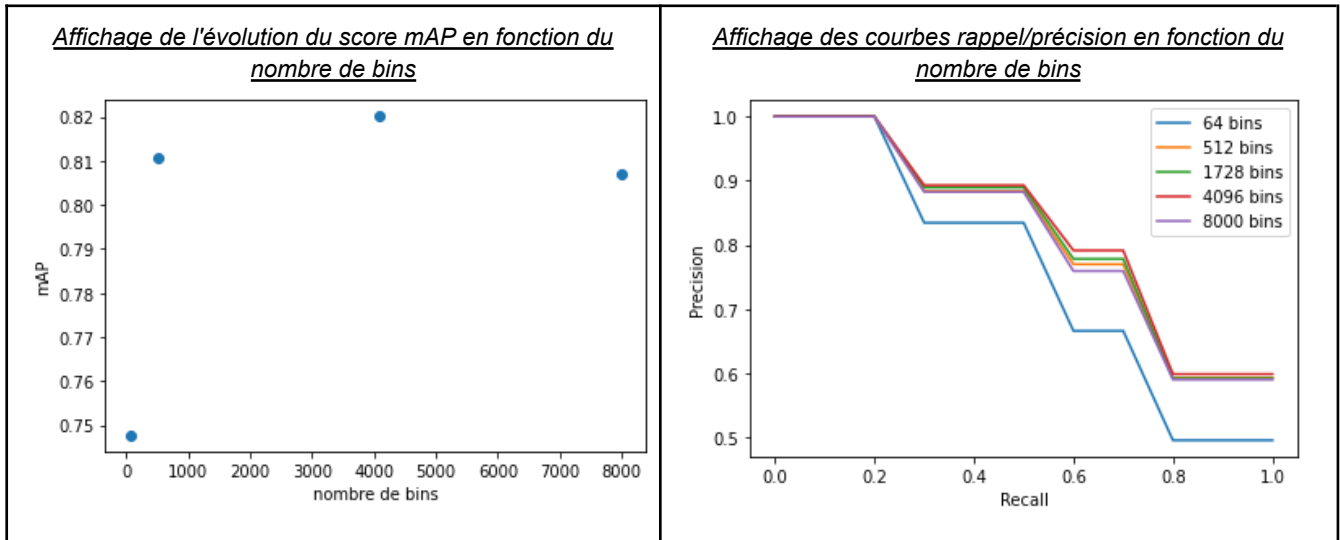
Score mAP de L1= 0.8137205411350266

Score mAP de L2= 0.7413455768690386

Confrontation des normes l1 l2 avec les courbes rappel/précision :

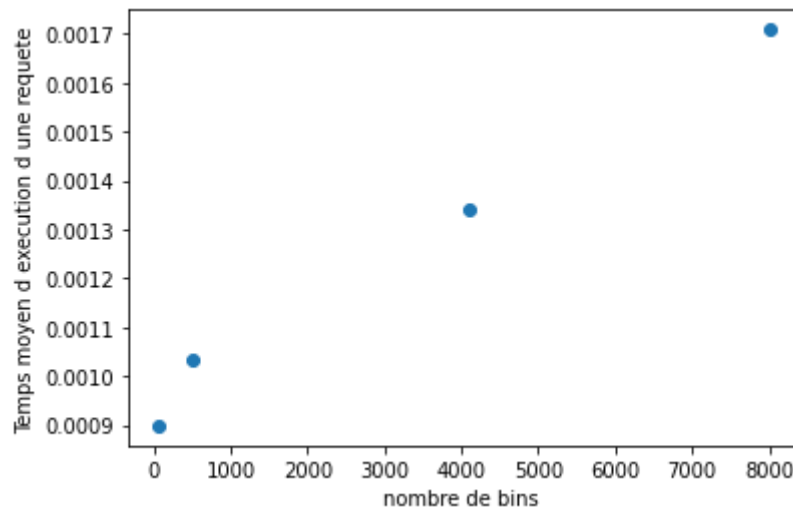


- b) Variation de la taille des "bins" : En faisant varier la taille des "bins", on observe que plus le "bin" est grand, meilleur est le score mAP jusqu'à un certain seuil. Pour trouver ce seuil optimal nous avons fait varier la taille [16, 512, 1728, 4096, 8000] "bins". Nous avons trouvé 4096 "bins" comme valeur optimale pour le score mAP et pour le critère Rappel-Précision. Aller au-delà de 4096 bins serait donc inutile car la précision n'en sera pas meilleure.



- c) Temps d'exécution d'une recherche : Nous pouvons observer une augmentation du temps moyen d'exécution d'une requête en fonction du nombre de "bins". Cela est dû à l'augmentation des dimensions des histogrammes, rendant les calculs de distances beaucoup plus chronophage. Cette évolution suit une forme linéaire.

Affichage de l'évolution du temps de calcul du requête en fonction du nombre de bin :



Pour conclure, nous observons que la distance de Minkowski semble optimale pour les systèmes par histogrammes. Pour notre base de données, 4096 de taille de "bins" semble fournir les meilleurs scores. Augmenter le nombre de "bins" ne sera d'aucune utilité car en plus d'obtenir une baisse de précision, le temps d'exécution sera plus conséquent étant donné l'augmentation du nombre de dimensions.

2) Sacs de mots visuels avec un index linéaire

- a) Variation des normes et de la normalisation : Nous avons fixé la taille du vocabulaire à 1000. Nous avons voulu analyser l'impact du changement de la norme pour calculer les distances entre les sacs de mots des images et l'intérêt de normaliser. Nous pouvons observer dans un premier temps l'intérêt de normaliser nos données en comparant le score mAP et les courbes Rappel-précision. Il y a une nette différence entre les scores normalisés et les non-normalisés. Il est nécessaire de normaliser les données car chaque image ne dispose pas du même nombre de descripteurs comme nous pouvons le voir sur l'image suivante qui représente la taille des 11 premières images. La première valeur du tableau correspond au nombre de descripteurs. Donc les valeurs au sein du vocabulaire ne sont pas comparables correctement sans normaliser au préalable. Dans un second temps, nous pouvons voir que lors du calcul des distances pour la recherche des requêtes les plus pertinentes L1 est plus performant.

Score mAP L1 normalisé = 0.86254093482541

Score mAP L2 normalisé = 0.7987041112701386

Score mAP L1 non-normalisé = 0.508189190824697

Score mAP L2 non-normalisé = 0.6746320723417735

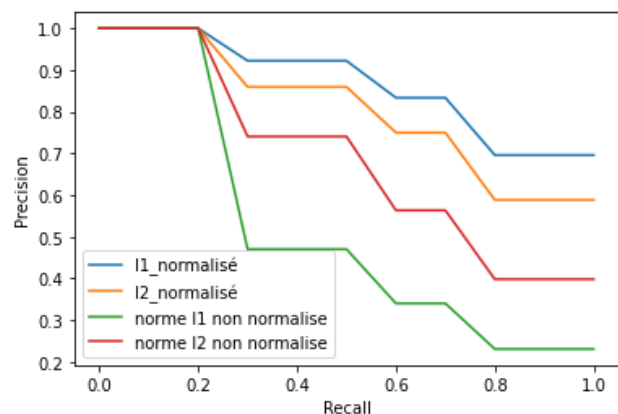
Nombre de descripteurs pour les 12 premières images

```

> describeurs = {dict: 500} {0:
> 0 = {ndarray: (669, 128)} [
> 1 = {ndarray: (836, 128)} [
> 2 = {ndarray: (964, 128)} [
> 3 = {ndarray: (1087, 128)} [
> 4 = {ndarray: (676, 128)} [
> 5 = {ndarray: (659, 128)} [
> 6 = {ndarray: (810, 128)} [
> 7 = {ndarray: (707, 128)} [
> 8 = {ndarray: (636, 128)} [
> 9 = {ndarray: (333, 128)} [
> 10 = {ndarray: (503, 128)} [
> 11 = {ndarray: (769, 128)} [

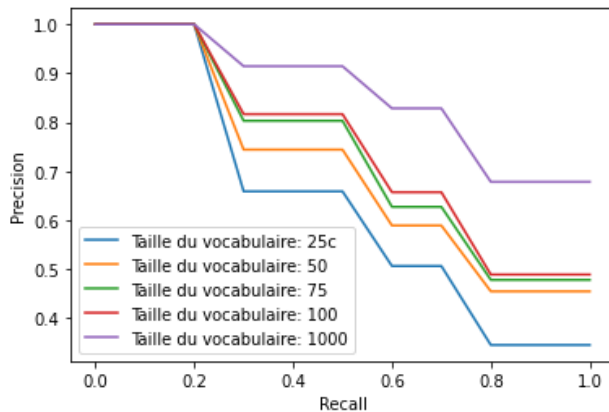
```

Affichage des courbes rappel précision en fonction de la norme choisie, pour les sacs de mots normalisés et non normalisés (Taille du vocabulaire: 1000 mots)

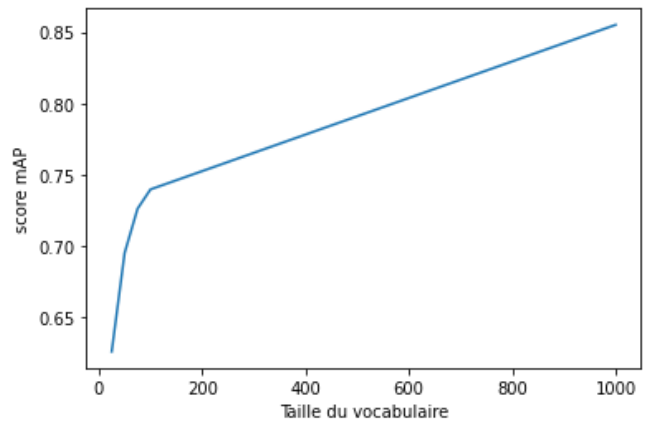


- b) Variation de la taille du vocabulaire : Nous avons voulu voir l'intérêt de la taille du vocabulaire. Nous avons d'abord compris que plus le vocabulaire est grand, meilleurs sont les résultats comme le montre les deux premiers graphiques.

courbes rappel précision en fonction de la taille du vocabulaire

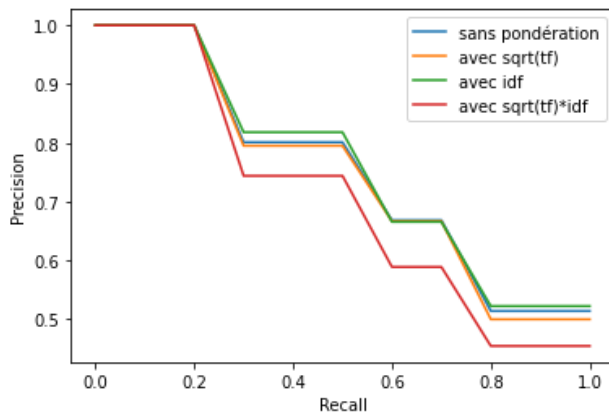


Evolution du score mAP en fonction de la taille du vocabulaire

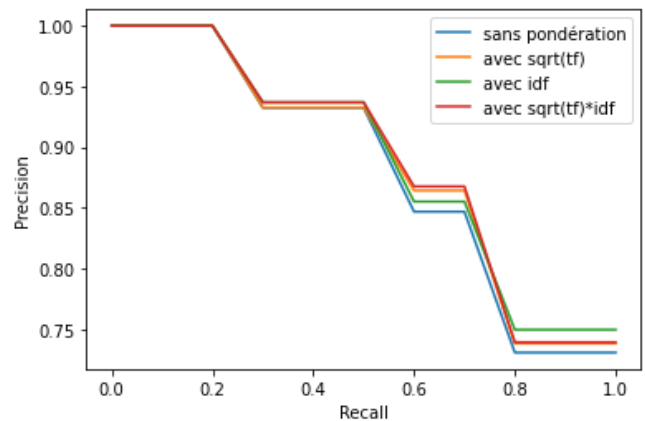


- c) Comparaison des pondérations entre deux vocabulaires : Nous avons comparé les scores des méthodes de pondérations entre deux tailles extrêmes du vocabulaire (50 et 5000). On peut voir que la pondération $\text{idf.racine}(\text{tf})$ ne fonctionne pas pour un vocabulaire de petite taille mais il permet d'avoir un meilleur résultat lorsque le vocabulaire croît.

courbes rappel précision des pondérations avec vocabulaire 50

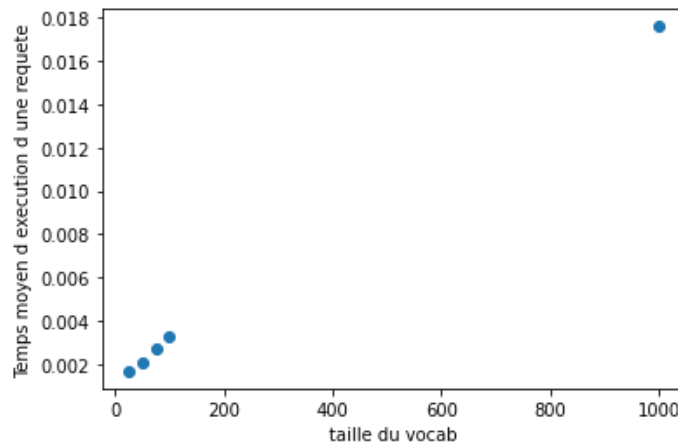


courbes rappel précision des pondérations avec vocabulaire 5000



- d) Temps de calcul en fonction de la taille du vocabulaire : Nous avons calculé le temps d'exécution d'une recherche pour un index linéaire. Nous voyons que le temps est dépendant du vocabulaire linéairement.

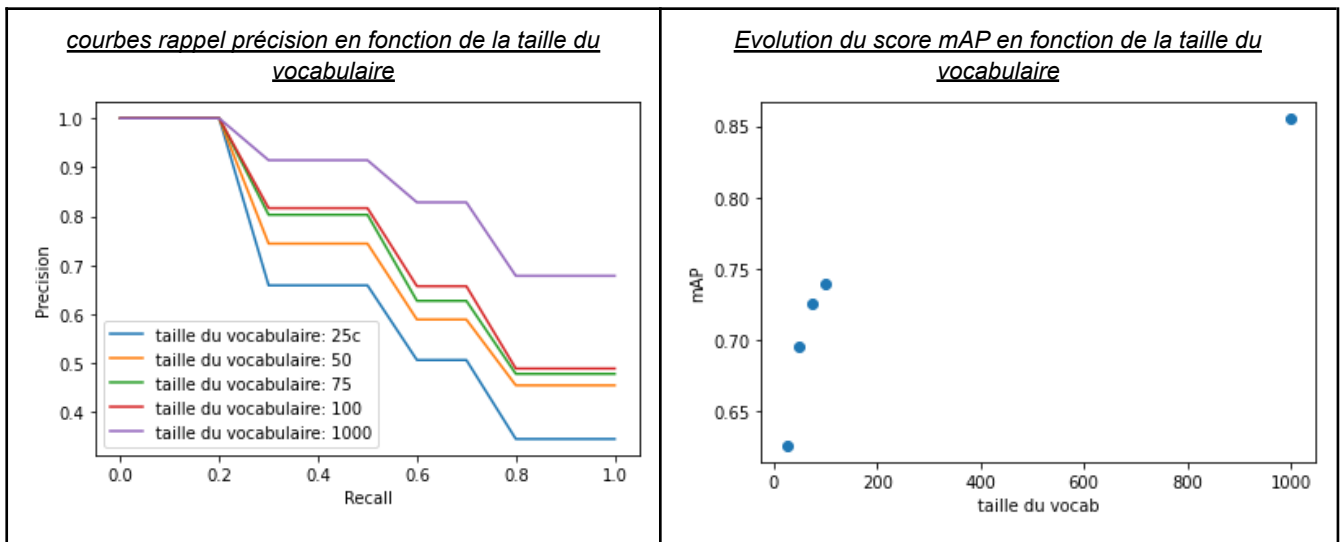
Temps de calcul moyen en fonction de la taille du vocabulaire



3) Sacs de mots visuels avec un index inversé

Pour l'analyse de l'index inversé nous n'étudierons que les paramètres concernant la taille du vocabulaire et le temps d'exécution car le principe du sac de mots reste le même. La seule différence vient de l'indexation des images.

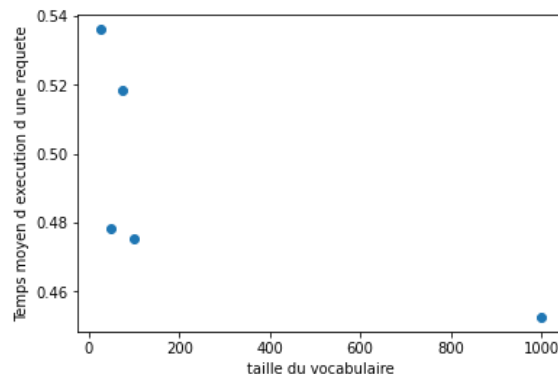
- a) Variation de la taille du vocabulaire : Nous pouvons voir que les scores obtenus sont très similaires à l'index linéaire. Nous pouvons donc dire que la taille du vocabulaire a le même impact sur les scores des recherches.



- b) Temps de calcul en fonction de la taille du vocabulaire : En observant le graphique, on comprend que le temps d'une recherche est supérieur à l'index linéaire. Cela pose problème car la recherche dans l'index inversé doit être théoriquement plus rapide que celle dans un index linéaire. La raison à un temps si élevé est que nous utilisons des tailles de vocabulaire trop faibles. Cela provoque une incidence sur la taille de la short-list à inspecter pour trouver l'image la plus proche. En effet, avec un vocabulaire de petite taille, la short-list se compose de tous les documents du corpus. L'établissement de la short-list représente alors une perte considérable de temps d'exécution. Augmenter la taille du vocabulaire pourrait résulter à des short-list

beaucoup plus petites et donc a un temps d'exécution inférieur à la recherche par index linéaire.

Temps de calcul moyen en fonction de la taille du vocabulaire

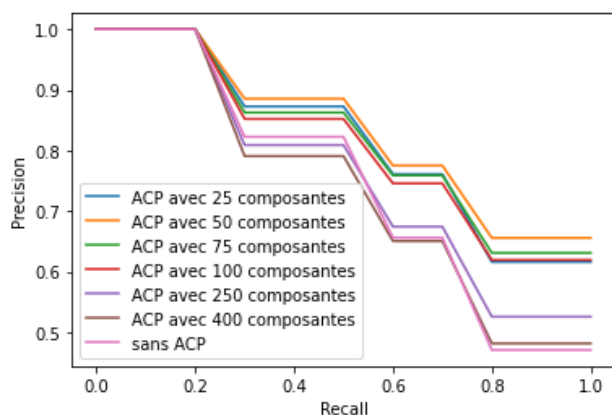


Pour conclure, nous pouvons remarquer que la norme L1 est la meilleure pour calculer les distances. Cette méthode est très intéressante et performante à partir du moment où le vocabulaire pour décrire une image est important. Cela impose un pré-traitement pour la création de l'index long. La recherche pour une requête n'est pas plus rapide que pour celle d'histogramme. Pour combler ce problème de temps, l'utilisation d'un index inversé permet à la méthode de sacs de mots de concurrencer les autres méthodes.

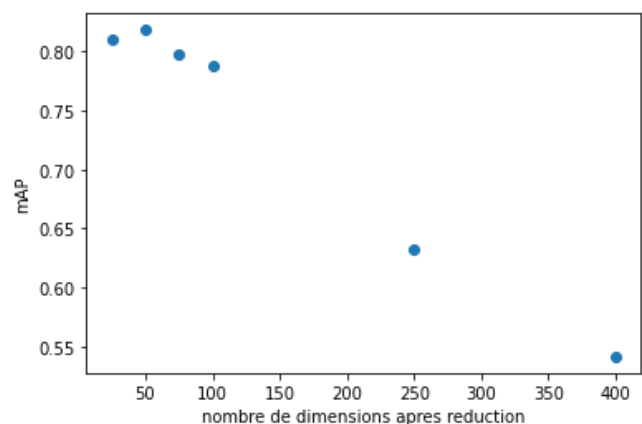
4) VLAD

- a) Variation des dimensions de l'ACP : L'ACP permet de diminuer le nombre de composantes et de mettre en valeur uniquement les informations pertinentes. À partir de ces graphes on peut observer que L'ACP participe à l'amélioration de la précision et du score mAP jusqu'à un certain seuil, où elle devient inefficace. Le système VLAD avec une ACP de 50 composantes est le plus optimal pour le score mAP. Aller au-delà de 50 composantes résulte à une perte de performances.

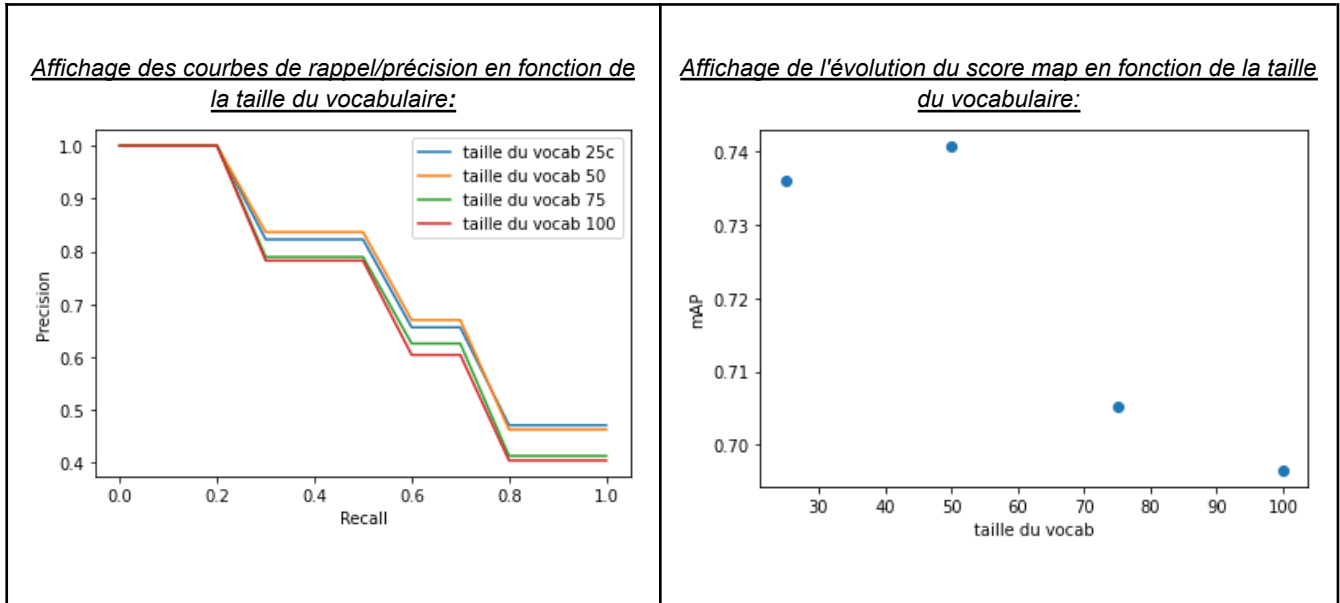
Affichage des courbes de rappel/précision en fonction du nombre de dimensions de l'ACP



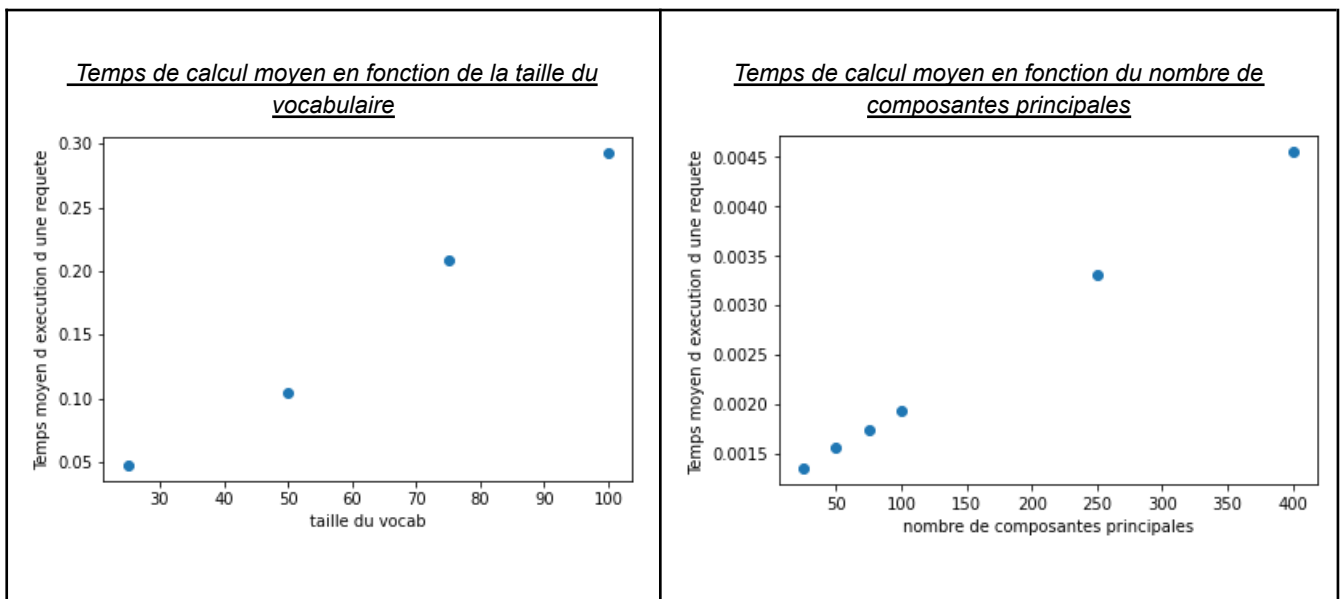
Affichage de l'évolution du score map en fonction du nombre de dimensions de l'ACP



- b) Variation de la taille de vocabulaire : Nous avons utilisé l'algorithme VLAD sans ACP afin de visualiser directement l'impact de la taille du vocabulaire sur les métriques. On peut donc constater que le descripteur VLAD fonctionne de manière plus précise avec un vocabulaire de petite taille. Sur les deux graphiques on observe que la taille optimale est de 50.



- c) Analyse du temps d'exécution: Nous avons pu remarquer que le temps évolue linéairement en fonction de nos deux critères. Lorsque l'on augmente le nombre de composantes, en plus de provoquer une perte de précision, le temps d'exécution augmente.



Pour conclure, nous pouvons remarquer que le système VLAD permet d'obtenir un très bon compromis précision/temps d'exécution. L'Analyse en composantes principales permet non seulement d'augmenter la précision, (à condition de bien choisir le nombre de composantes) mais de diminuer le temps d'exécution étant donné que le nombre de dimensions diminue. Un autre avantage du système VLAD est qu'il permet d'obtenir une bonne précision pour des vocabulaires de petite taille, puisqu'il permet d'avoir un maximum de précision pour une taille de vocabulaire de 50 mots

Conclusion Générale:

En faisant varier les paramètres de chaque système, nous avons pu analyser leurs forces et leurs faiblesses. Personnellement, ce TP/Projet nous a permis de mieux comprendre le fonctionnement de ces systèmes. Nous avons pu développer nos capacités d'analyse en termes d'optimisation de ces systèmes. Face à un problème de RI Image, nous pensons être maintenant capable de réaliser ce travail d'analyse en autonomie.