

Shallow representation - From Shallow to Deep representation for multimedia data - Lecture 6 : Image retrieval metrics

Johnny Nguyen¹

Abstract— This electronic document permits me to synthesis the sixth course of analysis and indexation.

I. INTRODUCTION

Nous avons une semaine pour faire le projet complet. Supprimer le softmax avec une ou deux couches. Pour récupérer l'image de la Tour Eiffel, nous devons combiner l'approche des descripteurs locaux et CNN. Plus le niveau se rapproche du softmax, plus nous nous rapprochons du résultat. Nous pouvons améliorer les données d'entraînement, l'architecture et refaire les données d'entraînement. L'architecture consiste à diviser une image en région et à en extraire un descripteur global. Le triplet "perte" consiste à avoir (la requête, pertinent, impertinent). Le SIFT est équivalent à la couche un et à la couche deux.

II. BACK TO BAG OF WORDS

L'index inversé d'un fichier correspond à admettre que chaque cluster de chaque image correspond à un mot. Savoir dans quels images se trouvent chaque cluster dans un tableau avec les clusters (mot) numérotés et les images numérotés. Le but est de diviser le nombre de d'occurrence d'un mot dans un document sur le nombre total de mot de ce document.

III. FOR THE SOUND

Pour l'analyse du son, l'architecture de Soundnet est disponible.

IV. CONCLUSION

La convolution est aussi appliquée sur les langues. Dans le lab nous allons comparé le deep learning à nos descripteurs locaux et globaux.

ACKNOWLEDGMENT

Thanks to Frederic Precioso for his work.

REFERENCES

- [1] <https://moodle.polytech.unice.fr/course/view.php?id=31>