

Initiation à la Recherche

Zied ARBI M2CIM DT

Titre	Text detection based on convolutional neural networks with spatial pyramid pooling.
Les Auteurs	R. Zhu, X J. Mao, Q H. Zhu, N. Li, Y B. Yang.
Publié	IEEE International Conference on Image Processing, classement B.
Le Thème	Détection et classification des textes apparaissant dans des images de scènes naturelles.
Le Problème	Plusieurs recherches ont traité ce sujet mais malheureusement il reste encore des obstacles à surmonter vu la diversité significative de ces textes à savoir les modèles, la complexité des informations d'arrière-plan ainsi les pouvoirs limités de MSER concernant la détection de composant non-textuel et également la taille de l'entrée de CNN qui doit être fixe ou compatible avec les couches entièrement connectées.
Les Théories et les Concepts	<ul style="list-style-type: none"> - SPP-net (Spatial Pyramid Pooling) : Technique qui permet aux réseaux de neurones de convolution (CNN) d'utiliser des images d'entrée de toute les tailles. - CNN (Convolutional Neural Networks) : Utilisés pour classer les images, les regrouper par similarité et effectuer la reconnaissance d'objet. - MSER (Maximally Stable Extremal Region) : Utilisées comme une méthode de détection des taches dans les images. - ANN (Artificial Neural Network) : c'est un paradigme de traitement de l'information inspiré par la manière dont les systèmes nerveux biologiques.
Méthode	<p>Avant de commencer à décrire la méthode proposée dans cet article il faut bien préciser que les méthodes existantes nécessitent souvent de rogner ou de déformer les images lors du prétraitement, ce qui réduit le pouvoir discriminant du CNN.</p> <p>En générale l'approche proposé dans cet article est basé sur deux axes fondamentaux : (Détection de textes et SPP-net pour la discrimination de texte).</p> <p>La première étape de cette méthode consiste à utiliser d'abord le MSER comme outil d'extraction des images d'entrée sous forme de propositions de caractères et afin d'améliorer le contraste de l'image 4 canaux (H, L, S et le gris) de MSER ont été décrits.</p> <p>La base de ce travail consiste à utiliser un nouveau descripteur de texte (SPP-net) pour le filtrage des faux caractères à partir des images d'entrée de toute les tailles et pas seulement celle de taille fixe. Le principe est le fait de remplacer la dernière couche de pooling par une couche de pooling pyramidale en générant des vecteurs de taille fixe en sortie, qui sont par la suite introduits dans la couche entièrement connectée.</p>

	<p>Enfin, en se basant sur les propriétés géométriques des caractères pour qu'ils puissent les regrouper dans des lignes de texte ainsi les diviser en mots.</p>
Les Résultats	<p>Les résultats de la méthode proposée ont été comparée aux celles de ICDAR 2011, 2013.</p> <p>Le contrôle de satisfaction de la qualité d'image (F-measure) est calculé à partir de deux paramètres qui sont : La Precision et Recall.</p> <p>La première évaluation de la performance de l'approche proposée dans cet article a été comparé avec cinq autres travaux en se basant dans un premier temps à l'ensemble de données ICDAR 2011.</p> <p><u>Tab1 (dans l'article)</u> : Les résultats obtenus sont respectivement 0,83, 0,76, 0,79 en Precision, Recall et F-measure, et qui montre une performance de F-measure d'un 1% de plus par rapport au meilleur résultat enregistré par la méthode de Huang.</p> <p>Ainsi les auteurs ont réalisé une deuxième évaluation de la performance de l'approche proposée en la comparant avec des autres méthodes en se basant ici sur l'ensemble de données ICDAR 2013.</p> <p><u>Tab2 (dans l'article)</u> : Une amélioration de F-measure de 2% à 3% a été enregistré par rapport à la méthode de Neumann.</p> <p>Les résultats obtenus ont été significativement concrètes en les comparant avec deux autres méthodes de discrimination de texte à savoir : MSER avec SVM et MSER avec CNN traditionnel <u>sans couche</u> de regroupement de pyramides spatiales plus précisément de 2%, et donc la conservation de la région de texte c'est-à-dire sans la déformer ou la recadrer reste la solution efficace.</p>
Critique	<p>Malgré les résultats significatifs de la méthode décrite dans cet article et pour une réalisation simple le réseau sur lequel ce travail est basé a été formé à partir des images à taille fixe.</p> <p>La possibilité de démarrer les calculs à partir des caractéristiques de textes qui sont recadrés ou bien redimensionnés pourraient aussi donner des résultats meilleurs, à confirmer dans les prochaines recherches.</p> <p>Les résultats obtenus ont été comparé à des autres qui remontent à des années un peu plus anciennes que la date de publication de cet article.</p> <p>A mon avis le traitement des données a été réalisé d'une manière centrale, je pense qu'aborder le sujet d'une façon distribué tel que le travail sur les systèmes multi-agents ou bien les micro-services va nous faire gagner du temps de calcul ainsi une facilitation de traitement de ce type de problème.</p>