Deep Structured Output Learning Dor Unconstrained Text Recognition

Li Honglin
Maxime Buron

16 février 2017

Table des matières

1 Résumé du papier

Nous avons choisi de travailler sur le papier intitulé Deep Structured Output Learning Dor Unconstrained Text Recognition de Max Jaderberg, Karen Simonnyan, Andrea Vedaldi et Andrew Zisserman. Voici un résumé de ce que nous en avons compris. L'article présente trois méthodes pour résoudre le problème suivant, dont la dernière combine les deux premières.

1.1 introduction du problème

Ce papier s'attaque au problème très général suivant : détecter et reconnaître du texte sur une image. Plus précisément, ce papier suppose la phase de détection est déjà réalisée, en d'autres termes l'on ne considère que des images contenant uniquement un seul mot sans autre contenu. Le sujet principal est donc la reconnaissance de mot, et particulièrement de mot sans contrainte, c'est à dire des mots qui ne sont pas obligatoirement issue d'une liste de vocabulaires.

1.2 encodage de l'entrée

Comme dit précedemment, l'entrée du problème est un image, seulement pour des questions de standardisation, notre algorithme nécessite des images de taille fixe 32×100 en noir et blanc. Malheuresement les images des ensembles de données ne sont pas toutes la même taille, c'est pourquoi elle sont étirées pour atteindre ces dimensions et cela sans préserver les proportions. Les images, une fois chargée sous forme de matrices de dimension 32×100 sont normalisées en leur soustrayant leur moyenne et en les divisant par leur écart type.

1.3 l'approche par caractère

L'approche par caractère considère que chaque caractère identifié selon sa position indépendemment des autres caractères.

La modélisation de la sortie de l'algorithme est une liste de N_{max} distribution de probabilités,où N_{max} est la longueur maximale qu'un mot sur une image en entrée. Les distribution de probalités modèlisent les lois de variables aléatoires à valeurs dans l'alphabet considèré pour identifier les mots augmenté par un caractère vide ϕ. Un mot ω composé des caractères $c_1c_2\ldots_n$ avec $n \leq N_{max}$ est encodé par la suite de N_{max} caractères suivantes $c_1, c_2, \ldots, c_n, \phi, \ldots, \phi$. La prédiction de l'algorithme pour une image donnée x est alors l'encodage ω^* défini par :

$$\omega^* = \arg\max_{\omega} P(w|x) = \arg \max_{c_1,c_2,\ldots,c_{N_{max}}} \prod_{i=1}^{N_{max}} P(c_i|\Phi(x))$$

où $\Phi(x)$ est un ensemble de paramètres.

L'algorithme du calcul des distributions est constitué d'un réseau de neurore conditionnel, qui sera réutiliser dans la

1.4 l'approche par sac de mots

1.5 l'approche jointe

1.6 les différents ensembles de données

2 La méthode choisie

3 L'implémentation

4 Les résultats

5 Conclusion