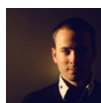


L'analyse du contenu des images par Google



Par Daniel Roch

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Fin 2014, Google a indiqué officiellement qu'il commençait à être capable de décrire le contenu d'une image de façon automatique, et de créer des descriptifs textuels visant à remplacer à terme les attributs ALT couramment utilisés à cette fin. Mais cette technologie est-elle assez fiable, au point d'être prochainement implémentée dans le moteur de recherche ?

Depuis longtemps, Google cherche à innover dans l'analyse des images et de leur contenu. En fin d'année dernière, le moteur de recherche a annoncé être capable d'analyser une image pour la décrire automatiquement. Nous allons voir ici en quoi consiste cette innovation ainsi que son impact actuel et potentiel sur le référencement naturel des images.

L'annonce de Google

Google l'a annoncé (<http://googlerecherche.blogspot.fr/2014/11/a-picture-is-worth-thousand-coherent.html>) : ses ingénieurs ont développé un système permettant d'analyser et de décrire automatiquement une image, comme le ferait le cerveau humain. L'objectif annoncé ici par Google est de pouvoir bien mieux comprendre de façon automatique chaque image qu'il indexe, pouvant potentiellement se passer des textes alternatifs de celles-ci (les fameux attributs ALT que l'on remplit manuellement).

Ainsi, Google serait capable dans l'image de la figure 1 de pouvoir générer la description suivante : "Two pizzas sitting on top of a stove top oven" (Deux pizzas posées sur une gazinière)

Le détail du brevet

Deux technologies sont utilisées ici :

- Un Recurrent Neural Network : le Deep Vision CNN ;



Fig 1. Une photo de deux pizzas sur un four

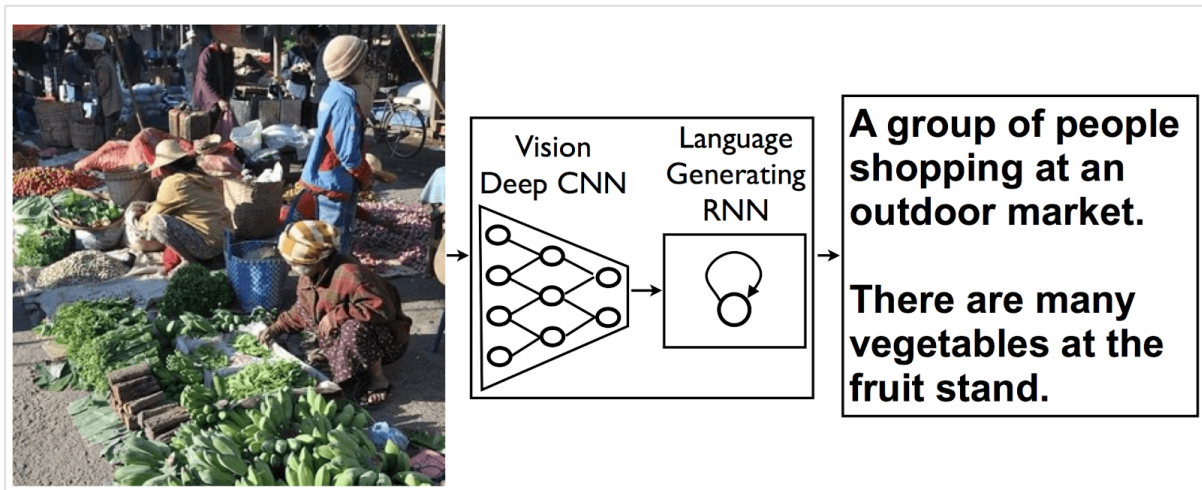


Fig 2. L'explication de la technologie fournie par Google

- Le *Language generating RNN* (pour générer le texte).

L'association de ces deux méthodes permettrait ainsi d'une part l'analyse de l'image, et d'autre part la génération d'une description humainement intelligible. Le brevet est détaillé ici, en anglais : <http://arxiv.org/abs/1411.4555>.

En réalité, Google a fait appel au *Machine Learning*, autrement dit des programmes informatiques qui apprennent au fur et mesure de leur travail. En d'autres termes, ce type de technologie apprendrait au fur et à mesure qu'il analyse des images. Pour mieux comprendre ce qu'est le *Machine Learning*, consultez cet article (en français) : <http://loic.knuchel.org/blog/2013/11/22/l-e-machine-learning-cest-quoi-exactement/>

Ce qui est important, ce n'est pas tant la technologie en elle-même, mais ce qu'elle peut impliquer sur le référencement naturel, comme nous le verrons un peu plus loin.

Google travaille depuis longtemps sur le sujet, notamment avec l'évènement ILSVRC, un challenge annuel sur la reconnaissance d'image et faciale, challenge pendant lequel Google s'est placé premier en 2014. Google parvenait déjà à l'époque à décomposer une même image en fonction de certains éléments qui la compose. L'exemple flagrant est l'image de la figure 3, fournie par la firme pour comprendre leur technologie (<http://googleresearch.blogspot.fr/2014/09/building-deeper-understanding-of-images.html>).

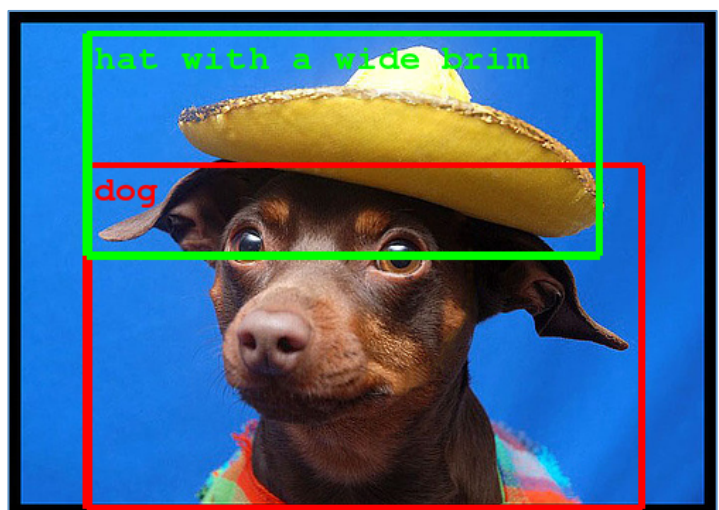


Fig 3. Google réussit ici à décomposer deux entités présentes dans l'image : d'une part le chapeau, et d'autre part le chien.



Fig 4. Une limace pris pour un serpent



Fig 5. Un poney pris pour un chien

Pas encore fiable

Dès 2013, Google pouvait déjà associer près de 1 100 termes à certaines photos et images.

La problématique est que la fiabilité n'est pas encore au rendez-vous. Lorsque le sujet se démarque du fond (comme dans l'exemple de l'image précédente) ou est uniforme, la technologie de Google permet au moteur de recherche de bien identifier chaque entité. Mais dans d'autres cas, c'est plus compliqué, comme dans les exemples des figures 4 et 5 donnés en 2013 (<http://googleresearch.blogspot.fr/2013/06/improving-photo-search-step-across.html>).

Pire encore, en réalité l'automatisation ne fonctionne pas à 100% de manière autonome. Il faut des « juges » humains pour valider les assertions du programme informatique. Dans les photos précédentes, l'algorithme s'était trompé. Si on le laisse faire, il risquerait à tort de cataloguer toutes les images de poney comme étant des chiens.

Les équipes de Google doivent ainsi pouvoir classer les différentes assertions du programme à l'aide d'un score BLEU (*Bilingual Evaluation Understudy*), c'est-à-dire la comparaison entre le rendu de la machine et celui d'un humain. On retrouve ainsi 4 niveaux de notation des descriptions des images dans notre cas de figure :

- Les descriptions sans erreurs ;
- Les descriptions avec erreurs mineures ;
- Les descriptions avec erreur importantes ;
- Les descriptions intégralement fausses.

Le programme pourrait même réussir à « inventer » des expressions inédites ou peu usuelles, créant de la confusion pour l'utilisateur final ou pour son propre algorithme. La machine apprend, mais Google doit valider toutes ses assertions et cadrer son évolution pour rester pertinent

(<http://fr.slideshare.net/frederickaplan/la-langue-comme-capital>).



Fig 6. Des exemples de notations BLEU de la technologie de Google

L'impact que ce brevet pourrait avoir

A terme, on pourrait ainsi imaginer pouvoir se passer intégralement du texte alternatif des images (attributs ALT) puisque les moteurs de recherche et les logiciels pourraient décrire de manière autonome le contenu réel de l'image analysée. Cela aurait d'ailleurs une portée plus large que les seuls moteurs de recherche, en incluant par exemple toutes les opportunités possibles en reconnaissance faciale ou en logiciel pour malvoyants.

Google pourrait réellement en tirer profit s'il parvenait à le rendre fiable et à l'implanter en temps réel dans son algorithme : cela éviterait de mauvaises analyses et classement d'images, et augmenterait de manière considérable la pertinence de ses résultats, lui permettant

ainsi de pouvoir fidéliser encore plus ses utilisateurs.

On pourrait ainsi éviter par exemple le résultat issu du test de l'agence Axenet, dans lequel Sylvain Richard est parvenu à positionner une image d'Adriana Karembeu sur le terme « Pizza aux anchois » (voir figure 7).

On pourrait aussi très bien imaginer que la recherche par image pourrait renvoyer à l'utilisateur une description beaucoup plus pertinente ou ciblée de l'image recherchée. Par exemple, lorsque vous êtes sur Google Image (https://images.google.com/?hl=fr&gws_rd=ssl), il est possible de glisser dans le champ de recherche n'importe quelle image pour effectuer une recherche web ou image par rapport au visuel fourni. On ne cherche plus par mot clé ou expression, mais par contenu (figure 8).

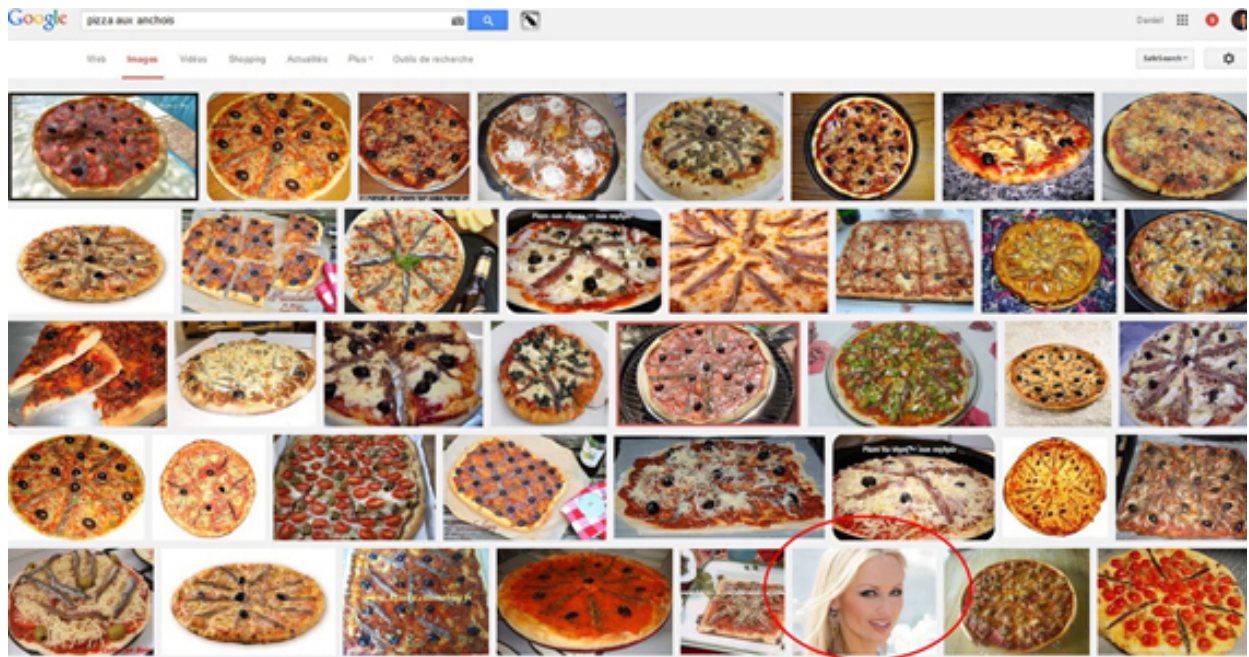


Fig 7. Une mauvaise analyse de Google sur le terme « Pizza aux anchois »
(<http://blog.axe-net.fr/google-images-pertinence-test-seo/>)



Fig 8. Dans cet exemple, nous glissons le logo d'Abondance dans la recherche Google Image



Fig 9. Google ne sait pas encore réellement analyser le contenu d'une image

Google va alors afficher une liste de résultats qu'il juge pertinents par rapport au visuel analysé (figure 9).

On constate alors très vite que la pertinence est très faible. Ici les premiers résultats n'ont aucun intérêt par rapport à l'image fournie, puisqu'il considère que le logo d'Abondance serait lié au terme « bonne année 2014 ».

L'implantation dans les moteurs de recherche

A l'heure actuelle, rien n'est fait. Comme le prouve les exemples précédents, Google ne dispose pas encore d'une technologie fiable pour pouvoir l'utiliser dans son algorithme de classement des sites.

Même si, depuis 2013, Google est sans doute parvenu à augmenter le nombre de termes qu'il sait associer automatiquement à une image (1 100 à l'époque),

on est très loin de la diversité du monde et des sites Web, surtout quand on sait qu'en 2014 nous avons :

- 3 milliards d'internautes ;
- 800 000 nouveaux sites Internet mis en ligne chaque jour ;
- 30 milliards de photos taguées par des utilisateurs de Facebook pendant l'année ;
- Et en 2012, 380 milliards de photos avaient été mises en ligne aux Etats-Unis.

Sources :

- <http://www.blogdumoderateur.com/chiffres-internet/> ;
- <http://www.planetoscope.com/developpement-durable/Internet->

Google doit donc d'abord réussir à consolider sa technologie avant de pouvoir l'utiliser.

Quand ce sera le cas, il se heurtera alors à un autre problème : les ressources serveur. Pour analyser les images, Google devra faire croître son parc de machines pour allouer des ressources à cette analyse, ressources dont Google doit déjà maîtriser la consommation.

On sait par exemple que Google est capable d'analyser en profondeur et en continu le contenu des sites, notamment en exécutant le Javascript. Cependant, il ne le fait pas constamment car cela consomme des ressources importantes, et il doit également faire des choix sur la pertinence ou non de venir analyser de nouveau une page présente dans son index.

Si cela vous intéresse, vous trouverez plus d'informations sur les serveurs de Google ici :

<http://www.google.com/intl/fr/about/datacenters/gallery/#/>

Conclusion

Il faut bien le reconnaître : avec cette annonce, Google a surtout fait de la communication pour prouver au monde entier son avancée dans le domaine, et le fait que l'entreprise dispose d'une technologie de pointe par rapport à ses concurrents.

Mais cette technologie, bien que prometteuse, est encore très loin d'être applicable en situation réelle. Pas assez fiable, nécessitant une validation humaine et consommant des ressources importantes, il faudra encore attendre au minimum quelques années avant de la voir éventuellement implantée dans le moteur de recherche de la firme.

Nous pouvons et devons donc continuer à remplir méticuleusement nos textes alternatifs ainsi que le contenu texte situé autour de nos images si nous souhaitons améliorer notre référencement naturel. En attendant mieux...



Daniel Roch, *Consultant*
WordPress, Référencement et
Webmarketing chez SeoMix
(<http://www.seomix.fr/>).