

Entrainement de YoloV3

Romain JAMINET

1^{er} avril 2019

Table des matières

1	Comment fonctionne Yolo	2
2	Plan d'action	2
3	L'entrainement	2
4	Amélioration	3

1 Comment fonctionne Yolo

Yolo fonctionne de la manière suivante :

1. On configure le réseau de neurone que nous allons utiliser (poids initiaux, couches, filtres...)
2. On donne la liste des objets que nous voulons qu'il détecte
3. Pour chaque image il faut créer un fichier texte contenant les informations suivantes :

— <Numéro du label> <centre x> <centre y> <largeur> <hauteur>

4. On démarre l'entraînement

2 Plan d'action

Pour entraîner Yolo à reconnaître des expressions faciales, j'ai utilisé un fichier de configuration qui est basé sur YoloV3 pour la reconnaissance d'objet.

J'ai réussi à obtenir une base de données d'image de 122Go soit environ un million d'image provenant en contactant l'université de Denver.

Il y a 11 labels différents : Neutre, Content, Triste, Surpris, Peur, Dégoût, Colère, Mépris, Rien, Incertain, Pas de visage, cependant il y a un problème, il faut que pour chaque image, je définisse le centre du rectangle du visage, ainsi que la largeur du rectangle et sa hauteur ce qui est vraiment dommage pour une base de données de cette taille.

Je vais donc devoir me limiter, je vais commencer par ne garder que quelques labels comme par exemple : Neutre, Content, Colère, Triste.

Je ne vais travailler que sur une cinquantaine d'image par label pour commencer soit environ 200 images. Je ne suis pas sûr que cela suffira pour la précision et la sûreté, il s'agit d'un essai.

Il faut également que je convertisse chaque image en .jpg et que je les redimensionne à la même taille. Vu que les images sont carrées, il n'y aura pas de déformation visuelle, juste éventuellement une perte de pixel qui n'est pas importante, car les images sont similaires en taille.

Les labels sont dans un fichier csv, je vais donc faire un script en python me permettant de les récupérer, comme je n'ai pas pu télécharger toutes les images, je vais simplement faire un script qui va agir de la sorte :

Pour chaque image dans le dossier des images :

1. Si l'image labélisé correspond à : content, triste, neutre, colère alors
 - (a) On créer un fichier texte nommé par le nom de l'image en .txt contenant le numéro du label
 - (b) Je vais utiliser Yolo en version entraîné pour détecter la position du visage et je vais rajouter à mon fichier les informations manquantes

Si la labélisation est rapide, je pourrais augmenter le nombre d'image.

3 L'entraînement

Après avoir labélisé 500 images à l'aide d'un script en python que j'ai développé, utilisant la librairie "face detector" (il s'agit d'une IA qui permet de tracer les rectangles autour des visages), j'ai par la suite déporté darknet sur Windows pour pouvoir l'utiliser avec ma carte graphique Nvidia, pour profiter des technologies CUDA et cuDNN. Le premier entraînement c'est donc effectué sur 500 images, ce qui est relativement peu, les développeurs de darknet conseillent environ 2000 données par classe donc il me faudrait environ 8000 images pour le moment avec les 4 classes (neutre, content, triste, colère) et 22000 si je veux utiliser tous les labels de la base de données (qui contient 1 million d'image).

Après avoir entraîné le réseau de neurones sur 500 images pendant environ 30 minutes, le taux de perte était inférieur à 1 donc c'était correct. J'ai donc fait des tests avec ce nouveau réseau de neurone entraîné, avec de nouvelles images et les résultats sont très correcte sur des images qui sont calibré dans les mêmes dimensions que les images d'entraînement, cependant le taux de certitude n'est pas assez élevé (environ 50%).

Cela est sûrement lié au temps d'apprentissage qui n'était pas très élevé et au nombre de données.

Le deuxième entraînement a été effectué avec un jeu de données de 2500 images, ce n'est toujours pas assez mais cela est tout de même 5 fois supérieur au premier jeu de données. Le taux de certitudes est désormais plus élevé

(environ 80%).

Pour avoir un taux proche de la perfection et une reconnaissance du visage instantanée, il faudrait que je diversifie mes données d'entraînement et que j'en augmente encore le nombre.

4 Amélioration

Bientôt