

Titre:

Détection de visages couplant template matching et apprentissage profond

Encadrant:

Julien Renoult (CNRS). Centre d'Ecologie Fonctionnelle et Evolutive de Montpellier (UM5175). julien.renault@cefe.cnrs.fr

William Puech (Univ. Montpellier). LIRMM. william.puech@lirmm.fr

Résumé:

L'objectif sera de développer une méthode par apprentissage profond permettant de détecter et de recadrer automatiquement des visages de singes sur photos. L'originalité du travail est double. 1) Le jeu de d'entraînement contient d'une part des photos originales montrant l'individu entier dans son environnement, d'autre part les portraits extraits manuellement des photos originales. 2) La détection devra être invariante à la rotation car les photos de portraits montrent des visages alignés, avec les yeux disposés sur un axe horizontal. Différentes méthodes pourront être explorées, à l'interface entre réseaux profonds et template matching.



Sujet détaillé :

La reconnaissance d'objets et de visages par réseaux de neurones profonds est un problème de la vision par ordinateur généralement considéré comme "résolu". Cependant, pour entraîner les algorithmes populaires tels que YOLO ou MASK_RCNN, il est nécessaire d'identifier les coordonnées du cadre délimitant la région d'intérêt (*bounding box*), à l'aide d'une application de type VGG Image Annotator. L'algorithme apprend à partir de l'image originale et d'un fichier renseignant ces coordonnées. Et si l'on dispose déjà d'images recadrées sur la région d'intérêt ? Peut-on exploiter ces images recadrées, ainsi que les images originales non recadrées directement pour apprendre un réseau à recadrer de nouvelles images ? Par ailleurs, l'image recadrée peut avoir été orientée, par exemple pour aligner les yeux sur un axe horizontal. Or, les algorithmes populaires de détection par apprentissage profond sont invariants à la rotation.

Les étudiants disposeront d'un jeu de données de quelques 15 000 photos non recadrées et des photos portraits correspondants pour explorer une ou plusieurs des trois voies de développement possibles : 1) utiliser une méthode classique et rapide de template matching (e.g., basée sur les SIFT) pour identifier les coordonnées du cadre ainsi que son angle de rotation, puis modifier un algorithme existant de détection par apprentissage

profond (e.g. YOLO) pour qu'il intègre l'information sur la rotation en plus de celle sur la taille et la position du cadre; 2) entraîner un réseau de neurones profond à aligner le visage sur l'image originale en apprenant l'angle de rotation (*deep regression*), puis utiliser SIFT pour apprendre les coordonnées du cadre, et enfin utiliser YOLO (dans sa version originale) pour détecter et recadrer les images; 3) développer un réseau de neurones apprenant directement sur les deux jeux de données (photos originales et recadrées) et donc intégrant le template matching à son apprentissage. Afin de comparer les différentes approches, nous évaluerons la similarité entre les images recadrées par le modèle et celles recadrées manuellement sur un jeu de données test.

Langage : Python

Mots clés :

face detection, bounding box, deep learning, template matching, rotation invariance

Références et liens pertinents:

- <https://github.com/toshi-k/kaggle-airbus-ship-detection-challenge>
- Liu, L., Pan, Z., & Lei, B. (2017). Learning a rotation invariant detector with rotatable bounding box. *arXiv preprint arXiv:1711.09405*.
- Li, S., Zhang, Z., Li, B., & Li, C. (2018). Multiscale rotated bounding Box-based deep learning method for detecting ship targets in remote sensing images. *Sensors*, 18(8), 2702.
- Sirmacek, B., & Unsalan, C. (2009). Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Transactions on Geoscience and Remote Sensing*, 47(4), 1156-1167.
- Zhang, Y., Zhang, Y., Shi, Z., Zhang, J., & Wei, M. (2019). Rotationally Unconstrained Region Proposals for Ship Target Segmentation in Optical Remote Sensing. *IEEE Access*, 7, 87049-87058.