Fiche de Lecture 4

Élodie Bouilleteau

Mercredi 15 Novembre 2018

1 Référence de l'article

Titre : REAL-TIME MULTIPLE PEOPLE TRACKING WITH DEEPLY LEARNED CANDIDATE SELECTION AND PERSON RE-IDENTIFICATION

Auteurs : Long Chen, Haizhou Ai, Zijie Zhuang et Chong Shang.

Université : Tsinghua University

Date de parution : 12 septembre 2018

Sujets : Computer Vision and Pattern Recognition.

Lien vers le document : `https://arxiv.org/abs/1809.04427v1`

2 Situation des auteurs

Long Chen, Haizhou Ai, Zijie Zhuang et Chong Shang sont 4 chercheurs du département informatique de l'université de Tsinghua.

3 Introduction

Cette publication traite d'une nouvelle approche de détection et de suivi d'une personne en temps réel. Ils proposent de détecter les détections de personne non fiable en collectant les candidats à partir des résultats de détection et de suivi d'une personne. Dans certaine situation, les système de détection et de suivi peuvent se compléter. D'une part, les détections fiables du traqueur peuvent être associé à court terme au détection en cas de détection manquante ou non précise. D'autre part, les résultats fiables des détections permettent d'éviter les écarts de détection des traqueurs. Afin de sélectionner en temps réel le candidat optimal, il utilise une nouvelle fonction de scoring basé sur un réseau de neurone convolutif.

4 Méthode de détection

Leur méthode de détection se base sur la sélection de candidat entre les méthodes de détection et de traçage.

Tout d'abord, ils mesurent tous les candidats à l'aide d'un score de notation unifié. Pour former cette fonction de score, ils fusionnent un classificateur d'objets entraîner de manière discriminatoire avec le degré de confiance du traqueur. Puis, à l'aide de la suppression non maximal effectuer sur les scores estimés, il obtiennent les candidats sans redondance.

Pendant la procédure d'entrainement de leur réseau, ils échantillonnaient aléatoirement des régions d'intérêts (candidat à classer avec les paramètres x0,y0,w,h) autour des vrais boîtes de détection et les considéraient comme des exemples positifs. Ils prenaient le même nombre de régions d'intérêts du fond de l'image (backgroon) comme des exemples négatifs. De cette manière, le réseau apprend à reconnaître les espaces spatial des objets.

5 Désigne du réseau

Leur classificateur se base sur le réseau de neurone entièrement convolutif basé sur une région de l'image (R-FCN). La carte des scores de l'image sont prédit en utilisant un réseau de neurone convolutif avec une architecture codeur-décodeur. La partie codeur est la couche légère centrale du réseau pour une performance en temps réel. Ils ajoutent la partie décodeur avec un sur-échantillonnage pour augmenter les cartes de scores des résolutions spatiales pour une classification future.

6 Méthode de traçage

La méthode de traçage consiste à prédire la localisation de chaque trace existante en utilisant le filtre de kalman. Ces prédictions sont adopter pour éviter les détections fausses causer par des variations visuelles des propriétés des objets ou par les occlusions dans les scènes de passage de personnes. Cependant, ces prédictions ne sont pas utilisable à long termes pour le traçage. Afin de mesurer le degré de confiance du filtre de kalman dont la précision peut décroitre s'il n'est pas mis à jour par détection au bout d'un certain temps, il utilise un indice de confiance par tracklet en utilisant des informations temporel.

Une trace peut être séparer en plusieurs tracklet.

7 Méthode d'association

Nous obtenons un candidat obtenu via la méthode de détection et un candidat obtenu via la méthode de traçage. Il vont utiliser la suppression non maximal pour sélectionner le candidat idéal.

8 Méthode de comparaison

Pour une meilleure précision, ils utilisent un réseau de neurone de comparaison des traits de caractères entre deux images afin de savoir s'il s'agit de la même personnes ou non.

9 Association hiérarchique des étapes de détection et de traçage

Premièrement, il appliquent l'association de données sur les candidats issus de la détection en utilisant l'apparence avec un seuil Td pour la distance maximal.

Ensuite, ils associent les candidats restant avec les traces non associé basé sur la jointure entre les candidats de détection et les candidats de traçages. Ils ne mettent à jour les représentation d'apparence que lorsque les candidats de traçage sont associer à une détection. La mise à jour est effectuer en sauvegardant les caractéristiques de ré identification (ReID) de la détection associé.

Ensuite, les nouvelles traces sont initialiser avec les résultats des détections restantes.

Avec l'association de données hiérarchique, ils ont seulement besoin d'extraire les fonctionnalités ReID pour les candidats de détection une fois par image. En combinant cela avec l'ancienne fonction de scoring efficace et les degrés de confiance des tracklet, leur infrastructure peut fonctionner à une vitesse en temps réel.

10 Performance

Ils utilisent SqueezeNet [16], la couche centrale de R-FCN pour la performance en temps réel. Leur réseau de neurone convolutif, composé de Squee-

zeNet et du décodeur, ne coûte que 8 ms pour estimer les cartes de scores pour une image d'entrée de la taille de 1152x640 sur un GPU GTX1080Ti.

Ils fixent k=7 pour les cartes de scores sensible à la position, et entraîne le réseau en utilisant l'optimiseur RMSprop avec un taux d'apprentissage de 1e-4 et une taille de lot de 32 pour 20 000 itérations.

Les données de formation pour la classification des personnes sont collectées à partir de MSCOCO.

11 Conclusion

Cette article explique une méthode en temps réel de détection de personne en utilisant une infrastructure composer de plusieurs partie : détection de personnes ressemblantes, vérification de détection des personne via l'association de candidat de détection et de traçage, utilisation de caractéristique de ré identification.

Ils traitent des détections non fiable en sélectionnant les candidats parmi les sorties de détection et de traçage.

La fonction de notation pour la sélection des candidats est formulée par un R-FCN efficace, qui partage les calculs sur toute l'image.

Ils améliorent la capacité d'identification des occlusions en introduisant les fonctionnalités ReID pour l'association des données.

Pour conclure, le traqueur proposer permet d'obtenir des résultats en temps réel et des performances de pointe sur la référence MOT16.