

Approche deep learning pour la prédiction de trajectoire de piétons à partir de données fisheye

Elisa-Souad Goulahsen*, Faten Chaieb-Chakchouk*,
Mohamed Djallel Dilmi*, Katarzyna Wolska*

* Efrei Research Lab, EFREI Paris-Panthéon-Assas Université,
30/32 Avenue de la République, 94800 Villejuif
elisa-souad.goulahsen@efrei.fr

1 Introduction

La prédiction de trajectoire de piétons permet d'anticiper leurs mouvements afin d'éviter les collisions et d'optimiser les espaces. La précision dépend des données visuelles utilisées. Les caméras fisheye offrent un champ de vision très large (180° contre 45° habituellement) (Schwalbe (2005)), permettent de capturer plus d'informations et couvrir des zones qui peuvent être exclues du champ de vision des caméras conventionnelles. Cependant l'exploitation de ces caméras reste limitée car elles introduisent des distorsions géométriques significatives dues à leur grand angle. Les méthodes adaptées exploitant les images fisheye sont insuffisantes, malgré leur potentiel. Les méthodes actuelles de prédiction sont adaptées aux images conventionnelles. Une méthode en particulier, Introvert (Shafiee et al. (2021)), une architecture sequence-to-sequence conçue pour prédire les trajectoires en tenant compte des informations dynamiques présentes dans la scène grâce à un mécanisme d'attention conditionnelle 3D. Une adaptation de cette approche repose sur l'intégration de la convolution déformable (Dai et al. (2017)) une technique permettant de rendre l'ajustement des filtres convolutifs dynamique en fonction des distorsions des images fisheye.

2 Travaux et recherches

Les distorsions résultent d'une projection non linéaire des rayons lumineux sur le plan image, entraînant une courbure des lignes droites contrairement aux caméras conventionnelles qui suivent une projection perspective linéaire sans distorsion. Cette déformation impact la précision dans la prédiction de trajectoires. Nous avons évalué l'impact des distorsions sur le modèle Introvert en mesurant les erreurs de déplacements moyens (ADE) et les erreurs de déplacement final (FDE). Les résultats (Table 1) montrent une dégradation des prédictions indiquant que le modèle actuel n'est pas adapté pour gérer les distorsions.

Pour remédier au problème, nous travaillons sur l'intégration de la convolution déformable dans Introvert afin de pouvoir tenir compte des distorsions. Contrairement à la convolution classique où les filtres ont des positions fixes appliquées uniformément sur toute l'image, la

Prédiction de trajectoire de piétons fisheye

	University	Zara1	Zara2	Hotel	ETH	AVG
Introvert baseline	0.21/0.33	0.16/0.26	0.16/0.25	0.15/0.74	0.10/0.15	0.21/0.35
Introvert fisheye	0.36/0.60	0.24/0.48	0.18/0.33	0.40/0.82	0.61/1.01	0.39/0.69

TAB. 1 – Comparaison de l’impact des distorsions fisheye sur les performances d’Introvert (architecture sequence-to-sequence) pour la prédiction de trajectoire : erreur de déplacement moyen (ADE) et final (FDE) sur différents datasets.

convolution déformable permet aux filtres de s’ajuster de manière dynamique en fonction des déformations présentes. Son efficacité a été démontré dans des tâches tel que la détection et la segmentation d’image avec l’intégration du modèle ResNet. Une amélioration notable a été observé dans la détection et la segmentation sémantique, comme le montre le Table 2 où la métrique mIoU (mean Intersection-over-Union) est utilisée pour évaluer la qualité, exprimée en pourcentage. Les informations dans ce tableau ont été extraites de l’article Dai et al. (2017) et mettent en évidence que l’ajustement des filtres rend le réseau capable de capturer de manière plus efficace les caractéristiques importantes et de mieux s’adapter aux distorsions présentes sur les formes des objets.

	ResNet-101(baseline)	ResNet-101+Deformable Conv
res5c	70.4	73.5
res5b,c	70.4	74.4
res5a,b,c	70.4	75.2
res5+res4b22,b21,b20	70.4	75.1

TAB. 2 – Resultats de l’ajout de convolutions déformables en mIoU (%) à différentes couches (res5c, res5b,c, res5a,b,c, res5 + res4b22,b21,b20) du réseau ResNet-101.

Dans cette étude nous montrerons les limites de la convolution classique sur les données fisheye et présenterons l’intégration de la convolution déformable dans les modèles actuels de prédiction de trajectoire de piétons. En effet, l’intégration de cette convolution permettrait de mieux capturer les informations visuelles en s’adaptant aux distorsions géométriques.

Références

- Dai, J., H. Qi, Y. Xiong, Y. Li, G. Zhang, et H. Hu (2017). Deformable convolutional networks. *2017 IEEE International Conference on Computer Vision*, 764–773, doi: 10.1109/ICCV.2017.89.
- Schwalbe, E. (2005). Geometric modelling and calibration of fisheye lens camera systems. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1–8.
- Shafiee, N., T. Padir, et E. Elhamifar (2021). Introvert : Human trajectory prediction via conditional 3d attention. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16815–16825, doi: 10.1109/CVPR46437.2021.01654.