

Segmentation d'instances de vêtements et métriques d'évaluation

Jouanneau, Warren, w.jouanneau@lectra.com, Lectra - LaBRI, Cestas (Orateur)

Bugeau, Aurélie, Univ. Bordeaux - Bordeaux INP - CNRS - LaBRI, Talence

Palyart, Marc, Lectra, Cestas

Papadakis, Nicolas, Univ. Bordeaux - CNRS – IMB, Talence

Veizard, Laurent, Lectra, Cestas

Résumé : **Thème 6 : Intelligence Artificielle**

Les images de vêtements jouent un rôle à tous les niveaux des chaînes de valeur des industries du textile et de la mode. Des problématiques d'accès et d'organisation des données émergent de l'abondance de ces images. Il est stratégique d'automatiser l'apposition de mots clefs permettant de les retrouver. Aussi, les méthodes de segmentation d'instances semblent proposer une avancée dans la localisation d'objets. Cette localisation permettrait de caractériser les vêtements contenus dans une image.

Mots clés : Vêtements et Mode, Vision par ordinateur, Segmentation d'instances, Métriques de discordance

1. Introduction

Pour l'industrie textile et de la mode, les images de vêtements ont une forte valeur à tous les niveaux du cycle de vie d'un produit. On les trouve par exemple, lors de la conception comme source d'inspiration, lors de séances d'essayage comme support de validation ou encore lors de la mise sur le marché comme référence visuelle. Il est donc nécessaire de faciliter leur accès et leur recherche parmi un grand nombre d'images candidates. Cela repose le plus souvent sur l'apposition manuelle de mots clefs afin de les indexer. L'automatisation de cette étape fastidieuse permettrait ainsi une économie considérable de temps et permettrait aux différents acteurs de se concentrer sur des tâches au cœur de leurs métiers.

2. Méthodologie

En apprentissage supervisé, cette problématique se réduit à une classification. Des travaux se sont tournés vers l'attribution d'une classe à une image entière par le biais de réseaux de neurones à convolution [1]. Cependant, cette approche donne de meilleurs résultats si un seul vêtement est présent. Afin de retrouver les classes de plusieurs vêtements présents, la classification peut s'opérer sur des sous-images n'en contenant qu'un seul [2] (*i.e.* détection), ou sur chacun des pixels de l'image [3] (*i.e.* segmentation sémantique). En couplant ces deux approches (*i.e.* segmentation d'instances), on obtient des masques qui contiennent les pixels d'un unique vêtement [4]. Ceci permet d'ajouter la distinction des instances d'une classe et une localisation plus fine à la détection. De plus, ces masques peuvent servir à la caractérisation des vêtements par d'autres méthodologies. Pour l'industrie, la segmentation d'instances est une étape cruciale pour l'indexation de contenu.

Parmi les corpus d'images de vêtements disponibles, peu incorporent les masques nécessaires à la segmentation d'instances (e.g. Modanet [5], iMaterialist [6]). Le corpus DeepFashion2 [7] est actuellement celui qui propose le plus grand nombre d'images et d'annotations avec masques.

Il est primordial de définir un protocole de validation et de test. Souvent, moins d'efforts sont fournis sur cette tâche que sur la définition de nouvelles méthodes de segmentation [8]. Or, l'évaluation a pour but de confronter un modèle au cas d'usage auquel il est supposé répondre. Le modèle peut alors être sanctionné par le biais de métriques. L'évaluation est critique à tous les niveaux du cycle de vie d'un modèle. Elle est critique, dans un premier temps, comme contrôle lors de l'entraînement, puis dans un deuxième temps, lors de la sélection du ou des modèles répondant au mieux à la tâche, et enfin comme détection de dérive en production.

3. Originalité / perspective

Lors de la constitution du corpus DeepFashion2, Mask R-CNN [9] a été évalué sur la segmentation d'instances. Les premiers travaux se sont donc tournés vers l'évaluation des méthodes de segmentation d'instances de l'état-de-l'art issues de Mask R-CNN appliquées à notre cas d'usage.

En segmentation d'instances la *mAP* est souvent utilisée comme métrique d'évaluation. Cette métrique issue du domaine de la recherche de document a subi plusieurs itérations et évolutions. Ces évolutions sont liées aux évaluations proposées par les corpus de l'état de l'art : Pascal VOC [10] et MS-COCO [11].

Les métriques utilisées pour la segmentation sémantique peuvent aussi servir pour évaluer la segmentation d'instances. Il existe une grande quantité de métriques et de nombreuses propositions d'ontologie les organisant (e.g. [12]). Cependant, il n'existe pas une métrique répondant à tous les contextes. Il est donc recommandable d'en sélectionner plusieurs évaluant différents critères [13]. On peut distinguer trois axes d'évaluation nécessaires à notre cas d'usage : - une évaluation globale des masques nous informant grossièrement sur la qualité des masques et reposant sur des métriques d'ensemble, - une évaluation des contours s'appuyant sur des métriques plus sensibles à localisation fine à proximité des contours, - une évaluation du contenu renseignant de la perte ou de l'ajout d'information par rapport aux masques de vérité terrain. Le but de cette dernière est d'évaluer a priori la faisabilité de la caractérisation.

Il sera aussi utile d'évaluer les méthodes dites en « une passe » (e.g. Yolact [14]). Cette évaluation pourra se faire au niveau système, en sanctionnant leur capacité à fournir des masques permettant la caractérisation des vêtements. Enfin, l'évaluation devra s'enrichir des métriques décrites selon les axes d'évaluation globale, de contour, de contenu. Ces étapes accomplies, les travaux pourront se tourner vers la caractérisation des vêtements. Dans un premier temps, les travaux se recentreront sur l'extraction de motifs tissu et l'extraction de couleur.

- [1] P. Gutierrez *et al.*, « Deep learning for automated tagging of fashion images », présenté à Europ. Conf. on Comp. Vis., 2018.
- [2] B. Lao et K. Jagadeesh, « Convolutional neural networks for fashion classification and object detection », présenté à Chinese Conf. on Comp. Vis., 2015.
- [3] W. Ji *et al.*, « Semantic Locality-Aware Deformable Network for Clothing Segmentation », présenté à Int. J. Conf. on Artif. Int., 2018.
- [4] W. Jouanneau, A. Bugeau, M. Palyart, N. Papadakis, et L. Vezard, « Étude comparative de méthodologies issues de Mask R-CNN : Application au Corpus DeepFashion2 », in *Reconnaissance des Formes, Image, Apprentissage et Perception*, Vannes, France, juin 2020, p. 1-3, Consulté le: déc. 21, 2020. [En ligne]. Disponible sur: <https://hal.archives-ouvertes.fr/hal-02649010>.
- [5] S. Zheng, F. Yang, M. H. Kiapour, et R. Piramuthu, « ModaNet: A Large-scale Street Fashion Dataset with Polygon Annotations », in *Proceedings of the 26th ACM international conference on Multimedia*, New York, NY, USA, oct. 2018, p. 1670–1678, doi: 10.1145/3240508.3240652.
- [6] S. Guo *et al.*, « The iMaterialist Fashion Attribute Dataset », présenté à IEEE/CVF International Conf. on Comp. Vis. Workshops, 2019.
- [7] Y. Ge, R. Zhang, X. Wang, X. Tang, et P. Luo, « Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images », présenté à IEEE Conf. on Comp. Vis. and Pat. Recogn., 2019.
- [8] Y. J. Zhang, « A survey on evaluation methods for image segmentation », *Pattern Recognit.*, vol. 29, n° 8, p. 1335–1346, 1996.
- [9] K. He, G. Gkioxari, P. Dollár, et R. Girshick, « Mask R-CNN », in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, p. 2980-2988, doi: 10.1109/ICCV.2017.322.
- [10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, et A. Zisserman, « The Pascal Visual Object Classes (VOC) Challenge », *Int. J. Comput. Vis.*, vol. 88, n° 2, p. 303–338, juin 2010.
- [11] T.-Y. Lin *et al.*, « Microsoft COCO: Common Objects in Context », in *Computer Vision – ECCV 2014*, Springer International Publishing, 2014, p. 740–755.
- [12] A. A. Taha et A. Hanbury, « Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool », *BMC Med. Imaging*, vol. 15, n° 1, p. 29, 2015.
- [13] P. Arbelaez, M. Maire, C. Fowlkes, et J. Malik, « Contour detection and hierarchical image segmentation », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, n° 5, p. 898–916, 2010.
- [14] D. Bolya, C. Zhou, F. Xiao, et Y. J. Lee, « YOLACT: Real-Time Instance Segmentation », in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), oct. 2019, p. 9156-9165, doi: 10.1109/ICCV.2019.00925.