



HAL
open science

Intégration des relations de composition / décomposition pour la segmentation d'objets en considérant les sous-parties et une classification hiérarchique

Christophe Pantel, Sylvie Chambon

► To cite this version:

Christophe Pantel, Sylvie Chambon. Intégration des relations de composition / décomposition pour la segmentation d'objets en considérant les sous-parties et une classification hiérarchique. *Reconnaissance des Formes, Images, Apprentissage et Perception*, Jul 2024, Lille, France. hal-04621321

HAL Id: hal-04621321

<https://hal.science/hal-04621321v1>

Submitted on 23 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intégration des relations de composition / décomposition pour la segmentation d'objets en considérant les sous-parties et une classification hiérarchique *

Christophe PANTEL

Sylvie CHAMBON

Université de Toulouse, IRIT, Toulouse INP, France

prenom.nom@toulouse-inp.fr

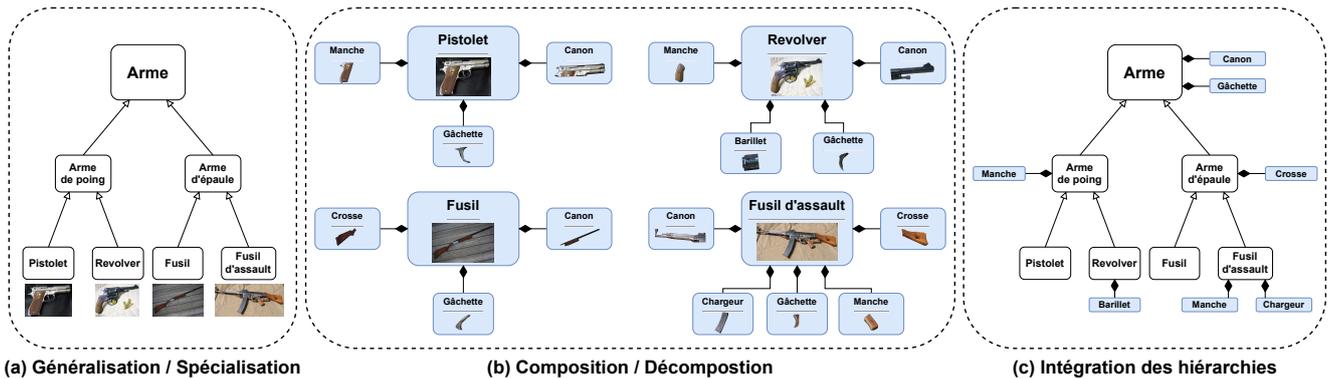


FIGURE 1 – Exemple de relations entre classes dans le cas des armes à feu (appelées armes dans la présentation)

1 Introduction à la segmentation hiérarchique

En vision, la segmentation identifie dans une scène des régions homogènes, cohérentes et dotées de sens qui peuvent être des objets élémentaires de la scène ou des parties d'objets plus complexes. Différentes architectures d'apprentissage profond sont utilisées pour cela comme FCN, *Fully Connected Networks* [16], les auto-encodeurs [1, 4] ou Mask-RCNN, *Region-based Convolutional Network* [11] ainsi que les mécanismes d'attention avec des Transformers [6, 18]. SAM, *Segment Anything* [13] est une approche de type *Foundational Models* : généralisation plus importante et spécialisation après apprentissage par *fine tuning* et *prompt engineering*. L'apprentissage minimise la différence entre sorties attendues et inférées, et les fonctions de perte, *loss function*, qui exploitent la sémantique des données pour choisir un minimum local plus pertinent.

Selon les objets étudiés, il reste difficile en segmentation d'identifier un objet composé de parties très différentes en apparence, par leur couleur, leur texture, la présence de problèmes liés à l'éclairage, les occultations ou encore le manque de contraste ou de différence avec le fond de la scène. Les armes de la figure 1 présentent un manche avec une couleur et une texture très différentes de celles du canon. L'objectif de nos travaux consiste à segmenter des objets afin de les classer. Par exemple, dans la figure 1, nous souhaitons distinguer un pistolet d'un revolver. Pour cela, nous proposons d'introduire une hiérarchie de classes, mais également une hiérarchie qui repose sur la décomposition en partie des objets (sur-segmentation en sous-parties de chaque objet). Par analogie avec l'analyse symbolique réalisée par l'être humain [2], nous formulons l'hypothèse que les sémantiques associées à ces différentes parties peuvent être combinées pour reconnaître et segmenter l'objet dans sa globalité. La segmentation par parties est principalement utilisée pour les humains [17] avec des architectures classiques [4] qui exploitent des jeux de données représentatifs des composantes d'objets comme PASCAL-Part [5] et PartImageNet [10]. En segmentation classique, une étiquette est attribuée à chaque point dans un ensemble de classes sémantiques disjointes. Les travaux de [8, 9, 12, 14, 15] améliorent la précision en exploitant une hiérarchie de classes et associent à chaque pixel plusieurs étiquettes suivant cette hiérarchie utilisée pour construire des fonctions de perte lors de l'apprentissage.

Nous proposons d'exploiter la classification issue de la modélisation symbolique des connaissances [3, 7] lors de la segmentation. Cette façon de procéder s'inspire du raisonnement humain. Nous faisons l'hypothèse que cela permettra d'améliorer la qualité de la segmentation à la fois au niveau des objets composites en exploitant l'identification de leurs composantes, et des objets composants à partir de l'identification des composites qui les contiennent. Les travaux de [14, 15] ont permis l'ajout des relations de spécialisation / généralisation et nous proposons de compléter cette modélisation en exploitant également les relations de composition / décomposition entre classes.

*Les travaux qui ont abouti à ces résultats ont été financés par le programme de recherche et d'innovation « Horizon Europe » de l'Union européenne dans le cadre de la subvention N° 101073876 (Ceasefire). Cette publication ne reflète que les opinions de l'auteur et de l'autrice. L'Union européenne n'est pas responsable de l'usage qui pourrait être fait des informations qui y sont contenues.

2 Intégration de la relation de composition / décomposition

La hiérarchie est définie comme une classification dans laquelle les termes sont ordonnés dans une relation de subordination, chaque terme dépendant du précédent et commandant le suivant. En représentation des connaissances, il est important de distinguer la hiérarchie des concepts selon :

- la **généralisation / spécialisation** issue de la classification des espèces en sciences du vivant (notion d'héritage en programmation objet) prise en compte par [14, 15];
- la **composition / décomposition**, que nous intégrons pour la segmentation hiérarchique des objets en parties.

Nous proposons d'intégrer ces deux hiérarchies illustrées par (a) et (b) en figure 1 pour obtenir (c). Par exemple, la sémantique arme de poing est la composition des sémantiques canon, gâchette et manche qui correspondent aux pièces composant une arme puis une arme de poing. Cette décomposition peut se répéter au niveau des composantes, appelées sous-régions, et conduire à une composition hiérarchique à partir d'éléments atomiques (éléments qui ne se décomposent plus).

Nous reprenons l'approche proposée dans [14] qui code la relation de généralisation / spécialisation comme des formules de logique des prédicats, traduites en fonction de pertes pour l'apprentissage en utilisant la logique floue. La table 1 rappelle les formules proposées dans ces travaux. La table 2 propose un codage pour les relations de composition / décomposition. Nous nous appuyons sur l'exemple de la figure 1 pour illustrer ces formules. Le terme \mathcal{C} désigne l'ensemble des classes du modèle de connaissance, $\mathcal{S} \subseteq \mathcal{C} \times \mathcal{C}$ désigne la relation de spécialisation et $\mathcal{D} \subseteq \mathcal{C} \times \mathcal{C}$ désigne la relation de décomposition. Nous notons $C_i \in \mathcal{C}$ une classe, $\{S_{i,j}\}$ ses sous-classes directes, $(C_i, S_{i,j}) \in \mathcal{S}$, et $\{P_{i,k}\}$ ses parties directes, $(C_i, P_{i,k}) \in \mathcal{D}$.

Règle et Principe	Formules	Exemple
Généralisation : Tout point d'une sous-classe $S_{i,j}$ de C_i est un point de C_i , $(C_i, S_{i,j}) \in \mathcal{S}$	$\forall p, S_{i,j}(p) \Rightarrow C_i(p)$	$\forall p, \text{Fusil}(p) \Rightarrow \text{Arme}(p)$
Spécialisation : Tout point de C_i est un point d'une de ses sous-classes $S_{i,j}$, $(C_i, S_{i,j}) \in \mathcal{S}$	$\forall p, C_i(p) \Rightarrow \bigvee_j S_{i,j}(p)$	$\forall p, \text{Arme}(p) \Rightarrow (\text{Poing}(p) \vee \text{Epaule}(p))$
Exclusion forte : Les sous-classes $S_{i,j}$ d'une classe C_i sont exclusives, $(C_i, S_{i,j}) \in \mathcal{S}$, $(C_i, S_{i,j'}) \in \mathcal{S}$	$j \neq j' \Rightarrow (\forall p, S_{i,j}(p) \Rightarrow \neg S_{i,j'}(p))$	$\forall p, \text{Fusil}(p) \Rightarrow \neg \text{Revolver}(p)$

TABLE 1 – Codage en logique de la relation de généralisation / spécialisation selon [14].

Règle et Principe	Formule	Exemple
Composition : Un point d'une classe composante $P_{i,k}$ de classes composites C_i est un point d'une des classes C_i	$\forall p, P_{i,k}(p) \Rightarrow \bigvee_i C_i(p)$	$\forall p, \text{Barillet}(p) \Rightarrow \text{Revolver}(p)$
Décomposition : Un point d'une classe composite C_i est un point d'une des classes composantes $P_{i,k}$ (si la décomposition est complète) $(C_i, P_{i,k}) \in \mathcal{D}$	$\forall p, C_i(p) \Rightarrow \bigvee_k P_{i,k}(p)$	$\forall p, \text{Arme}(p) \Rightarrow (\text{Canon}(p) \vee \text{Gachette}(p) \vee \text{Manche}(p))$
Exclusion forte : Un point d'une classe composite C_i ne peut appartenir qu'à une seule classe composante, $(C_i, P_{i,k}) \in \mathcal{D}$, $(C_i, P_{i,k'}) \in \mathcal{D}$	$k \neq k' \Rightarrow (\forall p, P_{i,k}(p) \Rightarrow \neg P_{i,k'}(p))$	$\forall p, \text{Canon}(p) \Rightarrow \neg \text{Manche}(p)$

TABLE 2 – Codage en logique de la relation de composition / décomposition.

Les différentes relations de (c) dans la figure 1 sont traduites en formules logiques combinées par une conjonction \wedge . Les exclusions fortes peuvent être relâchées pour prendre en compte la combinaison des relations (la formule associée plus complexe n'est pas détaillée). Les sous-parties contribuent à distinguer les classes et à identifier celles auxquelles appartient un même point. Les sommets de la hiérarchie contiennent les parties communes. En descendant dans la hiérarchie, les classes comportent des parties spécifiques et discriminantes. Par exemple, considérons un pistolet et un revolver, la présence du barillet permet de les distinguer. Par contre, considérons un pistolet et un fusil, seule la forme du manche, présent dans les deux armes, permet de les distinguer. De plus, si nous arrivons à distinguer une arme de poing d'une arme d'épaule, cette information, combinée à la forme spécifique du manche permet de distinguer un pistolet d'un fusil.

3 Perspectives

Nous proposons donc de coder les relations de composition / décomposition comme des formules de logique du premier ordre semblables à celles proposées par [14] pour coder les relations de généralisation / spécialisation. À chaque formule issue de ce codage est associée une fonction de perte qui prend en compte ces contraintes lors de l'apprentissage puis de l'inférence. Nous étudierons d'abord l'exploitation seule des relations de composition / décomposition pour valider notre proposition, puis la combinaison des deux familles de relations. Notre objectif est d'améliorer la qualité de la segmentation des objets composites.

Références

- [1] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 39(12) :2481–2495, 2017.
- [2] J. Bill, H. Pailian, S. J. Gershman, and J. Drugowitsch. Hierarchical structure is employed by humans during visual motion perception. Proceedings of the National Academy of Sciences, 117(39) :24581–24589, 2020.
- [3] R. Brachman and H. Levesque. Knowledge Representation and Reasoning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.
- [4] L.-C. Chen, I. Kokkinos, G. Papandreou, K. Murphy, and A. L. Yuille. DeepLab : Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 40(4) :834–848, 2018.
- [5] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. L. Yuille. Detect What You Can : Detecting and Representing Objects Using Holistic Models and Body Parts. IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pages 1979–1986, 2014.
- [6] B. Cheng, A. Schwing, and A. Kirillov. Per-pixel classification is not all you need for semantic segmentation. Advances in neural information processing systems, 34 :17864–17875, 2021.
- [7] R. Davis, H. E. Shrobe, and P. Szolovits. What Is a Knowledge Representation? AI Magazine, 14(1) :17–33, 1993.
- [8] E. Giunchiglia and T. Lukasiewicz. Multi-Label Classification Neural Networks with Hard Logical Constraints. ACM Journal of Artificial Intelligence Research, 72 :759–818, 2021.
- [9] E. Giunchiglia, M. C. Stoian, and T. Lukasiewicz. Deep Learning with Logical Constraints. In International Joint Conference on Artificial Intelligence, IJCAI, pages 5478–5485, 2022.
- [10] J. He, S. Yang, S., Y. Shaokang, A. Kortylewski, X. Yuan, J.-N. Chen, S. Liu, C. Yang, and A. Yuille. PartImageNet : A Large, High-Quality Dataset of Parts. arXiv, 2021.
- [11] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In IEEE International Conference on Computer Vision, ICCV, pages 2980–2988, 2017.
- [12] Z. Hu, X. Ma, Z. Liu, E. H. Hovy, and E. P. Xing. Harnessing Deep Neural Networks with Logic Rules. In Annual Meeting of the Association for computational Linguistics, ACL, 2016.
- [13] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick. Segment Anything. CoRR (arXiv), 2304.02643, 2023.
- [14] L. Li, W. Wang, and Y. Yang. LogicSeg : Parsing Visual Semantics with Neural Logic Learning and Reasoning. pages 4099–4110, 2023.
- [15] L. Li, T. Zhou, W. Wang, J. Li, and Y. Yang. Deep Hierarchical Semantic Segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pages 1236–1247, 2022.
- [16] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pages 3431–3440, jun 2015.
- [17] W. Wang, H. Zhu, J. Dai, Y. Pang, J. Shen, and L. Shao. Hierarchical Human Parsing With Typed Part-Relation Reasoning. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2020.
- [18] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. SegFormer : Simple and efficient design for semantic segmentation with transformers. Advances in neural information processing systems, 34 :12077–12090, 2021.