

Sélection d’algorithmes de classification par réseau de neurones

Joris GUERRY¹

Bertrand Le Saux¹

David Filliat²

¹ ONERA, Office national d’études et de recherches aérospatiales

² ENSTA, École nationale supérieure de techniques avancées

joris.guerry@onera.fr

Mots-clés

Fusion, Sélection d’algorithmes, Deep Learning, Réseaux de neurones, Classification.

Résumé

Pour permettre aux robots d’évoluer dans un environnement et d’y interagir il est nécessaire de leur permettre de reconnaître les objets qui les entourent. Pour cela la vision par ordinateur a massivement recours à l’apprentissage statistique (*machine learning*) qui consiste à entraîner le robot en lui présentant de très nombreuses images-exemples des objets à reconnaître. Ces dernières années les approches par réseaux de neurones ont permis d’améliorer significativement les résultats [1] [2] mais restent encore longues et délicates à entraîner et requièrent de grandes ressources de calcul. D’autre part, pour faire face à la haute variabilité des objets il est possible de recourir à des techniques de fusion [3]. Cette stratégie impose au minimum le coût de calcul cumulé de chaque méthode employée, alourdissant le processus, mais permet d’améliorer les performances de classification [4].

Nous avons donc cherché à contourner ce problème de temps et avons développé une solution pour sélectionner directement la méthode la plus propice pour une image donnée (figure 1).

Ce processus permet de réduire le coût de l’opération à celui du sélecteur de méthode plus celui de la méthode sélectionnée. Nous faisons l’hypothèse qu’il existe dans les images des caractéristiques discriminantes vis-à-vis des méthodes à employer. Pour notre étude nous nous sommes concentré sur un sous-ensemble du dataset ImageNet [2] et quatre méthodes à base de réseaux de neurones : SqueezeNet [5], GoogleNet [6], AlexNet [7], VGG [8].

AlexNet	100%			
GoogleNet	62.38%	100%		
SqueezeNet	61.95%	56.11%	100%	
VGG	63.38%	66.71%	57.08%	100%
	AlexNet	GoogleNet	SqueezeNet	VGG

TABLE 1 – Complémentarité des méthodes – % de Bonnes Classifications [BC] en commun entre deux méthodes.

Dans le tableau 1 la valeur de la case rouge signifie que seulement 56,11% = $\frac{\Omega(BC_{SqueezeNet} \cap BC_{GoogleNet})}{\Omega(BC_{SqueezeNet} \cup BC_{GoogleNet})}$ des images correctement classifiées par SqueezeNet ou GoogleNet l’ont été par les deux méthodes. Par déduction, 43,89% de ces images sont classifiées exclusivement par SqueezeNet ou par GoogleNet. C’est face à ces 43.89% d’images que le sélecteur est décisif.

Pour construire notre sélecteur nous avons opté pour un réseau de neurones basé sur l’architecture de SqueezeNet [5]. Ce réseau léger a pour objectif d’estimer la réussite que chaque méthode devrait avoir. Il apprend des *features* indépendantes des classes d’objets, qui doivent toutefois permettre de caractériser une certaine affinité des images avec les méthodes. Pour cela, nous définissons pour chaque image de l’ensemble d’entraînement un vecteur de réussite des méthodes qui servira de *label* (figure 2). En phase de test la meilleure sortie du réseau sélecteur indique la méthode à utiliser (figure 3).

Sur notre configuration PC faire tourner les quatre méthodes sélectionnées sur le dataset de test dure 151 secondes. Un oracle utilisant les différentes méthodes peut atteindre les 53,42% de réussite. Notre méthode permet d’atteindre 43,06% de réussite en seulement 76 secondes soit une amélioration en temps de calcul de 50%. En moyenne, le temps de calcul d’une méthode sur notre dataset est de 37 secondes et celui du sélecteur de 35 secondes. Notre méthode présente donc cet avantage temporel dès que le nombre de méthodes N_m est supérieur à 2 :

$$35s + 1 * 37s < N_m * 37s$$

Nous travaillons à produire une structure plus simple que SqueezeNet, spécialisée dans l’apprentissage de caractéristiques propres aux méthodes et espérons ainsi rattraper les performances d’une fusion.

Ces travaux s’inscrivent dans la réalisation d’un dataset multi-modal (image RVB, carte de profondeur et carte thermique) à l’ONERA et doivent permettre à un robot de choisir la méthode la plus adaptée aux données capteurs à un instant donné. Il serait ainsi envisageable de basculer d’une méthode basée couleur en extérieur à une méthode basée carte de profondeur en entrant dans un bâtiment non

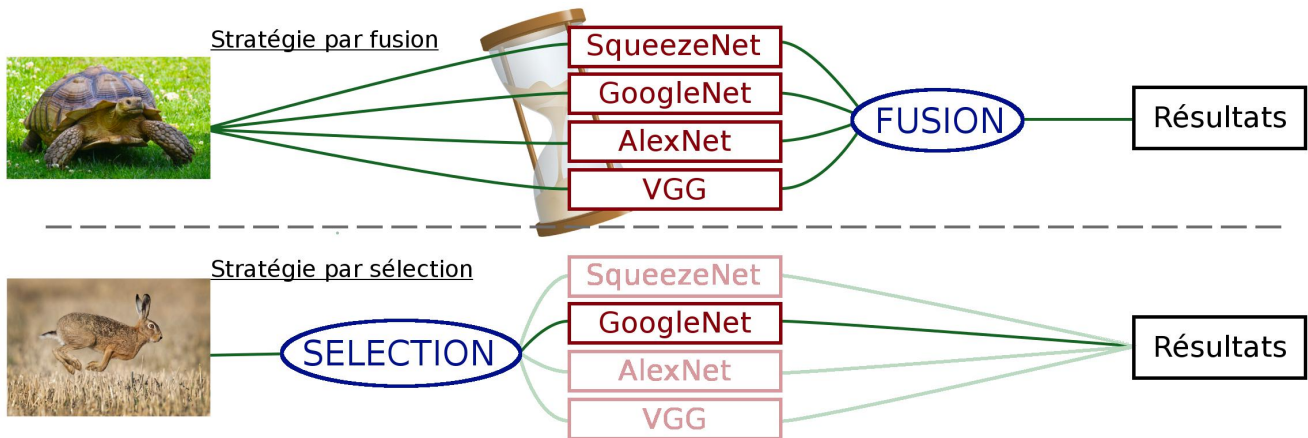


FIGURE 1 – Stratégie de classification par fusion VS. stratégie de classification par sélection. La force de la stratégie par sélection est de ne choisir qu’une seule méthode pour une image donnée en se basant sur des *features* de l’image d’entrée.

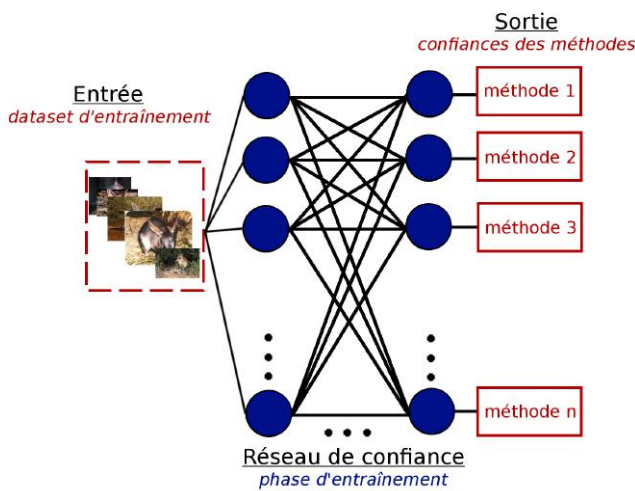


FIGURE 2 – Entraînement du réseau de confiance

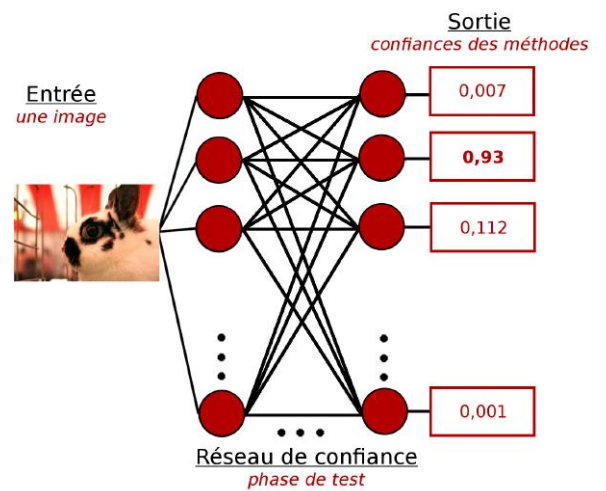


FIGURE 3 – Test du réseau de confiance

éclairé, jusqu’à ce qu’une information de chaleur soit captée et que le sélecteur propose d’utiliser une méthode de classification basée carte thermique...

Références

- [1] A. Krizhevsky and G. Hinton, “Learning multiple layers of features from tiny images,” 2009.
- [2] J. Deng, W. Dong, and R. Socher, “Imagenet : A large-scale hierarchical image database,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009*, 2009.
- [3] D. Lu and Q. Weng, “A survey of image classification methods and techniques for improving classification performance,” *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.
- [4] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, “Ensemble selection from libraries of models,” in *ICML*, 2004.
- [5] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, “Squeezenet : Alexnet-level accuracy with 50x fewer parameters and < 1mb model size,” *arXiv preprint arXiv :1602.07360*, 2016.
- [6] C. Szegedy, W. Liu, and Y. Jia, “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [8] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the devil in the details : Delving deep into convolutional nets,” in *British Machine Vision Conference*, 2014.