
Sujet de thèse - campagne 2020

Laboratoire : Institut des Systèmes Intelligents et de Robotique (ISIR)

Etablissement de rattachement : Sorbonne Université

Titre de la thèse : **Apprentissage de réseaux de neurones profonds légers et rapides : application à l'analyse faciale pour la robotique et l'interaction homme machine**

Directeur de thèse : Kevin BAILLY

Mail de contact : kevin.bailly@sorbonne-universite.fr

Codirection éventuelle :

Collaborations dans le cadre de la thèse :

Rattachement à un programme : ANR JCJC FacIL

Cotutelle envisagée :

Si oui avec quelle université & quel laboratoire :

Le sujet peut-il être publié sur le site web de l'ED SMAER : oui

Résumé du sujet :

Dans de nombreux domaines d'application et en particulier dans le domaine de la vision par ordinateur, les réseaux de neurones profonds ont obtenu des performances impressionnantes en comparaison à d'autres méthodes d'apprentissage statistique. Pour autant, les modèles appris sont souvent lourds et les temps d'inférence peuvent être longs. L'objectif de cette thèse est alors de proposer des modèles d'apprentissage profonds qui répondent aux contraintes d'empreinte mémoire et de temps de traitement inhérentes aux applications d'analyse faciale embarquées sur des systèmes interactifs (plateformes robotiques, agents conversationnels...). En particulier, le candidat explorera les méthodes de compression de réseau et proposera de nouvelles architectures dédiées légères et rapides.

ED SMAER (ED391)

Tour 45-46 Bureau 205- case courrier 270- 4, place Jussieu - 75252 PARIS Cedex 05

☎: 01 44 27 40 71

charlotte.vallin@sorbonne-universite.fr

Sujet développé

Contexte

Dans de nombreux domaines d'application et en particulier dans le domaine de la vision par ordinateur, les réseaux de neurones profonds ont obtenu des performances impressionnantes en comparaison à d'autres méthodes d'apprentissage statistique.

Pour autant, les modèles appris sont souvent lourds (plusieurs dizaines voire centaines de mégaoctets) et les temps d'inférence peuvent être longs, en particulier pour les architectures actuelles dont le nombre de couches ne cesse de croître (par exemple Resnet [He 2016] ou DenseNet [Huang 2017]). Ces contraintes restreignent ainsi leur exploitation dans des applications concrètes, en particulier dans le domaine de la robotique et des systèmes interactifs (ressources limitées, traitement temps réel)

Objectifs et résultats attendus

L'objectif de ce projet de thèse est de proposer des modèles d'apprentissage profonds qui répondent aux contraintes d'empreinte mémoire et de temps de traitement inhérentes aux applications d'analyse faciale embarquées sur des systèmes interactifs (plateformes robotiques, agents conversationnels...)

Dans la littérature, plusieurs stratégies ont été proposées que l'on peut répartir en deux catégories : la compression du réseau d'une part et la conception d'architectures dédiées d'autre part. La compression consiste à réduire la taille d'un réseau existant. Han et al. [2016] combinent des étapes d'élagage du réseau, de quantification de ses poids et de codage de Huffman, et montrent que cette stratégie réduit la taille d'un facteur 40 tout en conservant les capacités de prédiction du réseau. Des formes poussées de quantification, consistant à encoder les poids du réseau sous forme binaire [Hubara 2016] ou ternaire [Zhu 2017], ont aussi donné d'excellents résultats. Il est également possible de compresser l'information d'un réseau en apprenant des réseaux « légers » par imitation des prédictions d'un réseau « expert » [Hinton 2014, Bucilua 2006].

La seconde catégorie d'approches consiste à concevoir des architectures dédiées. On trouve dans la littérature, plusieurs architectures de réseau de neurones à convolution très efficaces dont certaines peuvent s'exécuter sur un smartphone [Howard 2017, Zhang 2017, Zoph 2017]. Ces méthodes s'appuient sur des convolutions séparables ou sur des opérations de convolutions groupées [Zhang 2017].

Après un état de l'art de ces approches, le candidat adaptera les méthodes les plus prometteuses au problème de l'analyse faciale temps réel (localisation de points caractéristiques, reconnaissance d'attributs...). Il pourra pour cela s'appuyer sur l'expertise et les méthodes développées par l'équipe depuis plus de dix ans.

Dans un second temps, le candidat explorera et proposera d'autres stratégies. Il est par exemple possible d'envisager des stratégies dynamiques de choix du réseau (ou de la sous-partie du réseau) qui sera explorée conditionnellement aux types d'images considérées ou d'informations sémantiques contextuelles de plus haut niveau.

Bibliographie

- [Han 2016] S. Han, H. Mao, and W. J. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. In ICLR, 2016
- [He 2016] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016
- [Howard 2017] A.G.Howard, M.Zhu, B.Chen, D.Kalenichenko, W.Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint, 2017.
- [Hubara 2016] Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio. Binarized neural networks. In NIPS, 2016.
- [Huang 2017] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In CVPR, 2017.
- [Zhu 2017] C Zhu, S Han, H Mao, WJ Dally, Trained Ternary Quantization? In ICLR, 2017.
- [Zhang 2017] X. Zhang, X. Zhou, M. Lin, and J. Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. arXiv preprint arXiv:1707.01083, 2017.
- [Zoph 2017] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le. Learning transferable architectures for scalable image recognition. arXiv preprint, 2017



**SORBONNE
UNIVERSITÉ**

**Ecole doctorale SMAER
Sciences Mécaniques, Acoustique, Electronique, Robotique**

Visa de la direction de l'unité

ED SMAER (ED391)

Tour 45-46 Bureau 205- case courrier 270- 4, place Jussieu - 75252 PARIS Cedex 05

☎: 01 44 27 40 71

charlotte.vallin@sorbonne-universite.fr