Accelerated AI for Edge Computing

Maryam Rahnemoonfar

Associate Professor of AI
Computer Vision and Remote Sensing Laboratory (Bina lab), UMBC
maryam@umbc.edu

1 Introduction

The volume and complexity of multidimensional signals have been growing exponentially. Analysis of the big datasets collected by various sensors remains a significant challenge for scientists and analysts. There is also a growing need for real-time analysis of data on Edge in many applications including emergency response, health care, surveillance, and cybersecurity. The faster one can harness insights from data, the greater the benefit in reducing costs, saving human lives, and increasing efficiency will be. While traditional analyses provide some insights into the data, the complexity, scale, and multi-disciplinary nature of the data necessitate advanced intelligent solutions. The success of recent data analytics technique based on deep learning have facilitated progress in a variety of tasks but they command significant computational complexity and require the availability of a large amount of labeled data for training.

2 Challenges and needs

Despite all recent advances of deep neural networks, there are currently several challenges; firstly, success of recent deep learning approaches for a variety of vision-based tasks highly depends on the availability of a large amount of labeled data for training. In many real-world applications, such as natural disasters and health-care, manually labeling images requires a significant amount of domain experts' time that could otherwise be spent on high-level scientific discovery. Secondly, most successful network architectures (such as VGG-net[1], Resnet[2]) have improved the performance of various vision tasks at the expense of significantly increased computational complexity and as a result, they need several days for training on graphics processing units

(GPUs). In many applications, fast analysis of data is vital. For example, after natural disasters, if one can reduce the initial response by one day, one can reduce the entire recovery by a thousand days [3]. Another example is in health-care; if one detects an elderly fall immediately, one can save his/her life. Thirdly, several studies have shown that even a limited amount of noise and perturbation greatly affects the performance of most deep learning techniques [4, 5]. Finally, the interpretability of black box representations of deep learning has long been the Achilles' heel of deep learning. It achieves superior performance at various tasks at the cost of low interpretability.

Critical need: There is a critical need in research community and industry to develop deep learning algorithms that are fast, unsupervised, less computational, and less sensitive to the noise.

3 Our vision

To address the aforementioned challenges in deep learning in general, we propose to devise new methodologies in time-frequency domain based on multi-resolution analysis. This domain-transformed deep learning approach has the potential to solve the issues of low training speed, and susceptibility to noise, as well as the need for large amounts of labeled data. Wavelets provide several advantages when performing deep learning operations in that domain, including sparse representation, multi-resolution, and space-frequency locality.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[3] Robin Murphy, "Ted talks: These robots come to the rescue after a disaster," 2015, https://www.youtube.com/watch?v=wG4RnDNWtJo, Last accessed on 2018-7-14.

[4] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*. IEEE, 2016, pp. 1–6.

[5] A. Fawzi, O. Fawzi, and P. Frossard, “Analysis of classifiers’ robustness to adversarial perturbations,” *Machine Learning*, vol. 107, no. 3, pp. 481–508, 2018.

[6] M. Rahnemoonfar, “Fast solution of phase unwrapping partial differential equation using wavelets,” *Electronic Journal of Differential Equations*, vol. 23, pp. 119–129, 2016.

[7] G. Beylkin, “On the representation of operators in bases of compactly supported wavelets,” *SIAM Journal on Numerical Analysis*, vol. 29, no. 6, pp. 1716–1740, 1992.

[8] G. Beylkin, R. Coifman, and V. Rokhlin, “Fast wavelet transforms and numerical algorithms i,” *Communications on pure and applied mathematics*, vol. 44, no. 2, pp. 141–183, 1991.

[9] I. Drori and D. Lischinski, “Fast multiresolution image operations in the wavelet domain,” *IEEE transactions on visualization and computer graphics*, vol. 9, no. 3, pp. 395–411, 2003.

[10] Y. Liu, M. M. Cheng, X. Hu, K. Wang, and X. Bai, “Richer convolutional features for edge detection,” in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on.* IEEE, 2017, pp. 5872–5881.

[11] J. Pont-Tuset, P. Arbelaez, J. T. Barron, F. Marques, and J. Malik, “Multiscale combinatorial grouping for image segmentation and object proposal generation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 1, pp. 128–140, 2017.

[12] T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-resolution image synthesis and semantic manipulation with conditional gans,” *arXiv preprint arXiv:1711.11585*, 2017.