ResNet-34: From research to Production

Scott Barnett, Alex Cummaudo, Rajesh Vasa

Applied Artificial Intelligence Institute
Deakin University
Melbourne, Victoria 3125, Australia

Abstract

Abstract goes in here

1. Introduction

Deep learning has taken off in recent years leading to the creation of many new models created by the research community. The field of computer vision in particular has gained significantly from deep learning [1, 2] especially in the areas of face recognition [3] and object detection [4]. While researchers continue to improve the models and provide new capabilities, industry is left to play catch-up struggling to find ways to include these new advancements in products and services. Additional work is often needed to transform academic research into tangible outcomes suitable for industry.

Objective To present experience from industry in incorporating state of the art Deep Learning models in a new product.

Methodology This paper presents a case study of the ResNet-34 Convolutional Neural Network from research paper to inclusion in Theia, a production face recognition system processing 1.2 million images. We accompany the case study with insights gained from our practical experience of dealing with a deep learning tookit [5] and the challenges with providing evidence that the black box algorithms work as stated in particular with analysing training datasets [6] and [7].

Contribution The two main contributions of this paper are (1) a set of challenges facing industry when incorporating state of the art technology in new products and (2) a proposal for a meta-model to facilitate the transfer of research to industry based on Empirical Software Engineering principles.

- A gap between the 'theoretical' research and implementing this into 'practical' product.
- The transition from research into an actual product.
- **Four-step process from pre-research in NN to the 'accessible' layer**: pre-research; NN architecture paper + initial implementation; commercial implementation (industry/production grade); accessible layer (DSL that is user friendly).
- Various roles for each 'step': Scientist/Researcher; Postdoc/PhD; Library Maintainer Engineer (Domain Specific); General Developer
- Thus, various skillsets needed. What are the gaps?
- In addition, need the model and data to make it work: need all three.

2. Related work

- **Case studies** of ML implementation from pure research.
- Look at Google, Facebook, MS Research.
- Look at the 'trace' of our **implemented work** in Theia.
- Trace back paper of FaceNet and see chain of prior research ("builds upon the work of XYZ") and see which initial implementations there were (Step 2) to see varying implementations.
- Could do the same for multi-task CNN.

3. Architecture

- Go into further detail about the 4 step process.
- What are the roles and why?
- What skillsets relate to each roles (i.e. (1) Scientist/Researcher = Theoretical/Heavily Mathematical; (2) Postdoc/PhD = Partially theoretical but can hack code together; (3) Library Maintainer Engineer = A little theoretical (knows concepts of domain); (4) General Developers = Practical.
- Important to gap these skills.

4. Model

- What data is used to train the model?
- Pre-processing steps needed to 'clean' the data. How does this affect the raw data itself?
- Additional models used to train this model?
- How can I update it? Is it locked down or can I use transfer learning?

5. Data

- How do I transform it?
- Hows does this influence the pre-processing stages in the model?
- How do you create data for more training?

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, IEEE, June 2016.

[2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, and T. Darrell, "Caffe - Convolutional Architecture for Fast Feature Embedding.," *ACM Multimedia*, 2014.

[3] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, IEEE, Mar. 2015.

[4] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. M. 0002, "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors.," *CVPR*, 2017.

[5] D. E. King, "Dlib-ml - A Machine Learning Toolkit.," *Journal of Machine Learning Research*, 2009.

[6] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *Procedings of the British Machine Vision Conference 2015*, pp. 41.1–41.12, British Machine Vision Association, 2015.

[7] H. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets.," *ICIP*, 2014.