Developing Deep Learning Models for Multimedia Applications in TensorFlow

Antonio José G. Busson
PUC-Rio
Rio de Janeiro, RJ, Brazil
busson@telemidia.puc-rio.br

Lucas Caracas de Figueiredo
PUC-Rio
Rio de Janeiro, RJ, Brazil
lucascf@tecgraf.puc-rio.br

Gabriel N. P. dos Santos
Faculdade ISL | Wyden
São Luis, MA, Brazil
gabrieainha@gmail.com

André Luiz de B. Damasceno
PUC-Rio
Rio de Janeiro, RJ, Brazil
andre@telemidia.puc-rio.br

Sérgio Colcher
PUC-Rio
Rio de Janeiro, RJ, Brazil
colcher@inf.puc-rio.br

Ruy Milidiú
PUC-Rio
Rio de Janeiro, RJ, Brazil
milidiu@inf.puc-rio.br

ABSTRACT

Methods based on Deep Learning became *state-of-the-art* in several Multimedia challenges. However, there is a gap of professionals to perform Deep Learning in the industry. Therefore, this short course aims to present the grounds and ways to develop multimedia applications using methods based on Deep Learning. Likewise, this short course is an opportunity for students and IT professionals can qualify yourselves.

CCS CONCEPTS

• **Applied computing → Interactive learning environments**; Digital libraries and archives;

KEYWORDS

Deep Learning, TensorFlow, Multimedia

ACM Reference format:
Antonio José G. Busson, Lucas Caracas de Figueiredo, Gabriel N. P. dos Santos, André Luiz de B. Damasceno, Sérgio Colcher, and Ruy Milidiú. 2018. Developing Deep Learning Models for Multimedia Applications in TensorFlow. In *Proceedings of WebMedia '18, Salvador-BA, Brazil, October 16–19, 2018,* 3 pages.
https://doi.org/https://doi.org/10.1145/3243082.3264605

1 INTRODUCTION

The disponibility of massive quantities of available data, alongside with the increasing computational capabilities, makes possible the development of more precise Machine Learning algorithms, providing advances in areas such as Natural Language Processing, Computer Vision and Audition. These results in new cognitive functionalities (e.g. learning, recognizing, detection) that can be used in multimedia applications, allowing the creation of mechanisms to be used in multimedia beyond the traditional use (e.g. capture, streaming and presentation).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
WebMedia '18, October 16–19, 2018, Salvador-BA, Brazil
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5867-5/18/10.
https://doi.org/https://doi.org/10.1145/3243082.3264605

Methods based on Deep Learning became *state-of-the-art* in several Multimedia challenges. This short course aims to present the grounds and ways to develop models using Deep Learning. Thus, this short course prepares the participant to: (1) understand and develop models based on Deep Neural Networks, Convolutional Neural Networks (CNN), Recurrent Neural Networks (e.g., LSTM and GRU); (2) apply the Deep Learning models to solve problems within the multimedia domain like Image Classification, Facial Recognition, Object Detection, Video Scenes Classification. The Python programming language is shown alongside TensorFlow, a package for developing Deep Learning models.

2 TUTORIAL ORGANIZATION

This short course is structured as follows:

(1) *Fundamentals of Deep Learning and introduction to Tensorflow (1 hour).* First, the minicourse presents basic concepts about machine learning: types of learning, tasks and datasets. Then, we show mathematical concepts for neural networks as: activation functions, loss functions and backpropagation algorithm. We also present a deep learning neural network using Tensorflow package.
(2) *Developing CNNs for image classification (1 hour).* We present concepts of Computer Vision and Convolutional Neural Networks (CNN). The basic operations of CNN (Convolutional and Pooling) are presented. After, we show three models of CNN: InceptionNet [5], ResNet [6], Inception-ResNet [4]. This step is concluded with the development of a CNN project to make images classification.
(3) *Facial recognition and object detection (1 hour and 15 minutes).* We present two models using CNNs to make tasks of facial recognition and objects detection using FaceNet [3] and YOLOv3 model [2], respectively to each task. These models have its implementation and execution performed using a specific dataset for both tasks.
(4) *Video classification (45 minutes).* Techniques of video classification are presented. First, we show techniques based on CNN to extract features in frames of videos. Then, these features are used in LSTM and GRU nets to perform the video classification. Finally, we show a implementation and execution of a video classifier using the YouTube8M dataset [1].

3 COURSE MOTIVATION TO TARGET PUBLIC

In the last years, Deep Learning allows a sharp advance in several areas of multimedia as speech process and computer vision. According to the last analyze of trend technologies, provided by Gartner institute [1], Deep Learning, Machine Learning and another related topics are on the top of the hype, indicating that such topics are one of the great expectations in the industry.

Due to this, we believe that this short course is interesting to students, researchers and IT professionals. Researchers, for example, can apply Deep Learning techniques to develop new projects of scientific research. Furthermore, there is a gap of professionals to perform Deep Learning in the industry. Therefore, this minicourse is an opportunity for students and IT professionals can qualify yourselves to this gap.

REFERENCES

[1] Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, and Sudheendra Vijayanarasimhan. 2016. Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675* (2016).

[2] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[3] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 815–823.

[4] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning.. In *AAAI*, Vol. 4. 12.

[5] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2818–2826.

[6] Sasha Targ, Diogo Almeida, and Kevin Lyman. 2016. Resnet in Resnet: generalizing residual architectures. *arXiv preprint arXiv:1603.08029* (2016).

BIO

MSc. Antonio Busson is Ph.D. candidate working under the guidance of Prof. Sergio Colcher at Pontifical Catholic University of Rio de Janeiro (PUC-Rio). He received a B.S. (2013) and M.S. (2015) in Computer Science from the Federal University of Maranhão (UFMA) on Brazil. He is researcher member at Telemidia Lab – PUC-Rio and research collaborator at the Laboratory of Advanced Web Systems (LAWS) – UFMA. His research interests are in multimedia systems, working mainly on the following topics: Coding and Processing of multimedia data; Hypermedia Document models, Pattern Recognition, and applications such as Web, iDTV and Games. Currently, He is working on the official Ginga Middleware development project, which is the middleware of the Japanese-Brazilian Digital TV System (ISDB-TB) and ITU-T H.761 recommendation for iPTV services. He also took part on the development of the GT-VOA in partnership with RNP, the National Research and Educational Network in Brazil, which was developed in the context of the RNP Working Groups program, during the cycles of 2012–2013, 2013–2014, and 2015. This work resulted in publications on conferences and symposiums such as CBIE 2014, SBGames 2016, WebMedia 2015 and 2016, and ACM Hypertext 2017. Address to access CV: http://lattes.cnpq.br/1857348479447184.

[1] https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/

MSc. Lucas Caracas de Figueiredo received a B.S. (2014) and M.S. (2017) in Computer Science from the Federal University of Maranhão (UFMA) on Brazil. Currently, he is Ph.D. candidate in Informatics at Pontifical Catholic University of Rio de Janeiro (PUC-Rio). His research interests include image processing, computer graphics and computational geometry. Address to access CV: http://lattes.cnpq.br/5353884964040661.

Gabriel Noronha Pereira dos Santos is B.S. candidate in Electrical Engineering at the University of Wyden on Brazil. His research interests include machine learning. Address to access CV: http://lattes.cnpq.br/8088572506239597.

MSc. André Damasceno received a B.S. (2013) and M.S. (2015) in Computer Science from the Federal University of Maranhão (UFMA) on Brazil. Currently, he is Ph.D. candidate in Informatics at Pontifical Catholic University of Rio de Janeiro (PUC-Rio). His research interests include multimedia systems and data science, working mainly on the following topics: educational data mining and learning analytics. Address to access CV: http://lattes.cnpq.br/0969337931297570.

Ph.D. Sérgio Colcher received B.S. (1991) in Computer Engineer, M.S. (1993) in Computer Science and Ph.D. (1999) in Informatics, all by PUC-Rio, in addition to the postdoctoral (2003) at ISIMA (Institute Supérieur d'Informatique et de Modelisation des Applications - Université Blaise Pascal, Clermont Ferrand, France). Currently,

he teaches in Informatics Department at Pontifical Catholic University of Rio de Janeiro (PUC-Rio). His research interests include computer networks, analysis of performance of computer systems, multimedia/hypermedia systems and Digital TV. Address to access CV: http://lattes.cnpq.br/1104157433492666.

Ph.D. Ruy Luiz Milidiú received B.S. (1974) in Mathematics and M.S. (1978) in Applied Mathematics from Federal University of Rio de Janeiro. He also received a M.S. (1983) and Ph.D. (1985) in Operations Research from University of California. Currently, he teaches in Informatics Department at Pontifical Catholic University of Rio de Janeiro (PUC-Rio). He has experience in Computer Science, focusing on Algorithmics, Machine Learning and Computational Complexity. Address to access CV: http://lattes.cnpq.br/6918010504362643.