Tutorial--MSWCAS 2021

Title: Recurrent Neural Networks (RNNs) Intricacies: from simple to gated architectures

Summary/abstract:
This 2-hour tutorial begins in the 1st hour with the basic & foundation of architectures of simple and gated (e.g., LSTM and GRU, and also Slim Gated) RNNs. It also covers the fundamentals of the core (stochastic) gradient descent and its principled connection to non-convex optimization with constraints. It shows how recurrent networks extend feedforward networks and how their operation is distinct from memory-based "Associative Memory" feedback neural nets. In the 2nd hour, the tutorial uses (Google's) Cloud Computing frameworks and Tensorflow-Keras to demo varied families of RNNs performance. The tutorial is suitable for researchers, engineers, data scientists, and students who want a deeper understanding of RNNs for formulating challenging engineering opportunities with deep learning and accelerated cloud computing. It introduces key nuances of (recurrent) neural networks, deep learning, benchmark examples, and learning strategies to effectively solve engineering challenges in (image) classification and detection. Extensions to other media modalities are itemized to enable applicability to text, speech, and audio signals.

Instructor: Professor Fathi M. Salem
Circuits, Systems, and Neural Networks Laboratory
Dept. of Electrical and Computer Engineering/ University Neuroscience Program,
Michigan State University.
East Lansing, Michigan 48824-1226 (USA)

Dr. Salem has extensive experience with Google and Nvidia Cloud Computing Platforms.

Prerequisites: Familiarity with basic linear algebra and statics as well as programming fundamentals such as functions and variables

Deep Learning UI/Frameworks: any of Keras, Tensorflow, Caffe, DIGITS.

Duration: 2 hours

Tutorial Description:

Learn the basics of deep learning by (hands-on) training deep neural networks on the Google Cloud Frameworks--using CPU and GPUs-- to improve performance and capabilities. We use image datasets to demonstrate the training and inference capabilities; however, the learned methodologies apply to other data modalities from text, to speech, to images, and videos. Subsequent strategies for trained networks (model) deployment are explored.

Specifically, the tutorial will include:

(i) Implement benchmark deep learning architectures and workflows for applications in image classification and object detection

(ii) Effectively exploring (hyper-) parameters, network structure, and other strategies to improve and increase (accuracy) performance and capability

(iii) Approaches to deploy your (trained) neural networks onto an embedded platform to start solving real-world problems

List of References:

For Cloud Computational Platforms:

[1] keras.io (https://keras.io/ online source for using the IU Keras with Tensorflow as a backend) or Tensorflow.keras.

For technical and theoretical details:

[1] DeepLearning.org (free textbook, available online).
[2] Atra Akandeh and Fathi M. Salem, "Slim LSTM networks: LSTM_6 and LSTM_C6," (available on arXiv.org: **arXiv:1901.06401** [**pdf**, **other**])
[3] Daniel Kent and Fathi M. Salem, "Performance of Three Slim Variants of The Long Short-Term Memory (LSTM) Layer," (available on arXiv.org: **arXiv:1901.00525**)
[4] Fathi M. Salem, "Slim LSTMs," (available on arXiv.org: **arXiv:1812.11391**)
[5] Joel C. Heck and Fathi M. Salem, "Simplified minimal gated unit variations for recurrent neural networks," in the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, August 2017, pp. 1593 – 1596.
[6] Rahul Dey and Fathi M. Salem, "Gate-variants of Gated Recurrent Unit (GRU) neural networks," in the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), August 2017, Boston, MA, August 2017, pp. 1597 – 1600.
[7] Yuzhen Lu and Fathi M. Salem, "Simplified gating in long short-term memory (LSTM) recurrent neural networks, " 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), August 2017, pp. 1601 – 1604.
[8] F. M. Salem, "A Basic Recurrent Neural Network Model," December 2016, available via arXiv.org as arXiv:1612.09022.
[9] F. M. Salem, Y. Wang, and R. Y. Choi, "On the Analysis of Dynamic Feedback Neural Nets," IEEE Trans. on Circuits and Systems, Vol. 38, No. 2, February 1991.
[10] F. M. Salem and S. Bai, "A New Feedback Neural Network with Supervised Learning," IEEE Transactions on Neural Networks, 2 (1), January 1991, pp. 170-173.

Brief biography of the speaker:

Dr. Salem has extensive experience with Google and Nvidia Cloud Computing Platforms.

Dr. Salem's current research interests include: Neural Networks and Learning Systems, Blind Signal Deconvolution and Extraction, Dynamical Systems and Chaos, Integrated CMOS Sensing and Processing. He was the Chairman of the CAS Technical Committee on Neural Systems and Their Applications (1997–1998). He served on the IEEE Neural Network Council (1999–2000), and was the first Vice President of the IEEE Neural Network Council for Technical Activities (1999–2001). He was the Guest Co-Editor of the IEEE-CAS Special Issue on Bifurcations and Chaos in Circuits and Systems July 1988 (with T. Matsumoto), and the Special Issue on Micro-Electronic Hardware Implementation of Soft Computing: Neural and Fuzzy Networks with Learning, Journal of Computers and Electrical Engineering, July 1999 (with T. Yamakawa). He was the recipient of the IEEE CAS Golden Jubilee Award (1999), the IEEE Third Millennium Award (2000), and The CAS Darlington Best Paper Award (2001). He was an Associate Editor and Guest Editor for numerous IEEE and other transactions including the IEEE Circuits and Systems, IEEE Neural Networks, the Journal of Circuits, Systems, and Computers, and the Journal of Computer and Electrical Engineering. He served in several capacities in several conferences including the General Chair of the IEEE Midwest Symposium on Circuits and Systems (MWSCAS) in Lansing, MI, in 2000 (and will be the MWSCAS General Chair in 2021).

He has authored more than 300 technical papers, and co-edited the textbook (*Dynamical Systems Approaches to Nonlinear Problems in Circuits and Systems*, (SIAM, 1988). He is a co-inventor of more than 14 patents on adaptive nonlinear signal processing, neural networks, and sensors. He was a Distinguished Lecturer of the IEEE CAS Society in 2000–2001. He has developed and delivered more than 20 tutorials at conferences and universities.