COLLEGE OF ENGINEERING AND COMPUTER SCIENCE
FLORIDA ATLANTIC UNIVERSITY

Announces the Ph.D. Dissertation Defense of

Connor Shorten

for the degree of Doctor of Philosophy (Ph.D.)

“Data Augmentation in Deep Learning”

DATE May 1st, 2023, 10 AM
Virtual Dissertation

Zoom
Meeting ID: 845 1153 9336
Passcode: 629340

DEPARTMENT:
Electrical Engineering and Computer Science

ADVISOR:
Taghi M. Khoshgoftaar, Ph.D.

PH.D. SUPERVISORY COMMITTEE:
Taghi M. Khoshgoftaar, Ph.D., Chair
Borko Furht, Ph.D.
Xingquan Zhu, Ph.D.
DingDing Wang, Ph.D.
Mehrdad Nojoumian, Ph.D.

ABSTRACT OF DISSERTATION

Data Augmentation in Deep Learning

Recent successes of Deep Learning-powered AI are largely due to the trio of: algorithms, GPU computing, and big data. Data could take the shape of hospital records, satellite images, or the text in this paragraph. Deep Learning algorithms typically need massive collections of data before they can make reliable predictions. This limitation inspired investigation into a class of techniques referred to as Data Augmentation. Data Augmentation was originally developed as a set of label-preserving transformations used in order to simulate large datasets from small ones. For example, imagine developing a classifier that categorizes images as either a “cat” or a “dog”. After initial collection and labeling, there may only be 500 of these images, which are not enough data points to train a Deep Learning model. By transforming these images with Data Augmentations such as rotations and brightness modifications,more labeled images are available for model training and classification. In addition to applications for learning from limited labeled data, Data Augmentation can also be used for generalization testing. For example, we can augment the test set to set the visual style of images to “winter” and see how that impacts the performance of a stop sign detector.

The dissertation begins with an overview of Deep Learning methods such as neural network architectures, gradient descent optimization, and generalization testing. Following an initial description of this technology, the dissertation explains overfitting. Overfitting is the crux of Deep Learning methods in which improvements to the training set do not lead to improvements on the testing set. To the rescue are Data Augmentation techniques, of which the dissertation presents an overview of the augmentations used for both image and text data, as well as the promising potential of generative data augmentation with models such as ChatGPT. The dissertation then describes three major experimental works revolving around CIFAR-10 image classification, language modeling a novel dataset of Keras information, and patient survival classification from COVID-19

Electronic Health Records. The dissertation concludes with a reflection on the evolution of limitations of Deep Learning and directions for future work.

BIOGRAPHICAL SKETCH

Born in Connecticut, USA
B.S., Florida Atlantic University, Boca Raton, Florida 2018
M.S., Florida Atlantic University, Boca Raton, Florida 2019
Ph.D., Florida Atlantic University, Boca Raton, Florida 2023

CONCERNING PERIOD OF PREPARATION
& QUALIFYING EXAMINATION

Time in Preparation: 2019 - 2023

Qualifying Examination Passed: Fall 2020

Published Papers:

C. Shorten and T. M. Khoshgoftaar. “A survey on image data augmentation for deep learning”. Journal of Big Data, 6(1):48, 2019.

C. Shorten, T. M. Khoshgoftaar, and B. Furht. “Deep Learning applications for COVID-19”. Journal of Big Data, 8(1):54. 2021.

C. Shorten, T. M. Khoshgoftaar, and B. Furht. “Text Data Augmentation for Deep Learning”. Journal of Big Data 8(1):34, 2021.

C.Shorten and T. M. Khoshgoftaar. “KerasBERT: modeling the Keras language”. In 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 219-226. IEEE, 2021.

C. Shorten and T.M. Khoshgoftaar. “Investigating the Generalization of Image Classifiers with Augmented Test Sets”. In 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI), pages 10-17. IEEE, 2021.

C. Shorten, E. Cardenas, T. M. Khoshgoftaar, J. Hashemi, S. G. Dalmida, D. Newman, D. Datta, L. Martinez, C. Sareli, and P. Eckard. “Exploring Language-Interfaced Fine-Tuning for COVID-19 Patient Survival Classification”. In 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), 2022.

C. Shorten and T. M. Khoshgoftaar. “Language Models for Deep Learning Programming: A Case Study with Keras”. In Deep Learning Applications, Volume 4, 135-161, 2022.

C. Shorten and T. M. Khoshgoftaar. “An Exploration of Consistency Learning with Data Augmentation”. In The International FLAIRS (Florida Artificial Intelligence Research Society) Conference Proceedings 35, 2022.

E. Cardenas, C. Shorten, T.M. Khoshgoftaar, and B. Furht. “A Comparison of House Price Classification with Structured and Unstructured Text Data”. In The International FLAIRS (Florida Artificial Intelligence Research Society) Conference Proceedings 35, 2022.

C. Shorten, T. M. Khoshgoftaar, J. Hashemi, S. G. Dalmida, D. Newman, D. Datta, L. Martinez, C. Sareli, and P. Eckard. “Predicting the Severity of COVID-19 Respiratory Illness with Deep Learning”. In The International FLAIRS (Florida Artificial Intelligence Research Society) Conference Proceedings 35, 2022.