



**Deep Learning Program  
Triangle Machine Learning Day  
September 20, 2019**

**SPEAKER TITLES/ABSTRACT**

**Xipeng Shen**

North Carolina State University

“Adaptive Deep Reuse for Deep Learning”

The speed of Deep Neural Networks (DNN), in both training and inference, is important for its practical usage. This talk presents adaptive deep reuse, a novel optimization to enhance the speed of DNN by efficiently and effectively identifying unnecessary computations in DNN training on the fly. By avoiding these computations, the technique cuts the training time of DNN by 69% and inference time by 50%, with virtually no accuracy loss. The method is fully automatic and ready to be adopted, requiring neither manual code changes nor extra computing resource. It offers a promising way to substantially reduce both the time and energy cost in both the development and deployment of AI products. Since its recent publication, the technique has drawn a lot of interest in media, industry practitioners, and research community.