A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks

E. M. Dogo, O. J. Afolabi, N. I. Nwulu, B. Twala, C.O. Aigbavboa

In this paper, we perform a comparative evaluation of seven most commonly used first-order stochastic gradient-based optimization techniques in a simple Convolutional Neural Network (ConvNet) architectural setup. The investigated techniques are the Stochastic Gradient Descent (SGD), with vanilla (vSGD), with momentum (SGDm), with momentum and nesterov (SGDm+n)), Root Mean Square Propagation (RMSProp), Adaptive Moment Estimation (Adam), Adaptive Gradient (AdaGrad), Adaptive Delta (AdaDelta), Adaptive moment estimation Extension based on infinity norm (Adamax) and Nesterov-accelerated Adaptive Moment Estimation (Nadam). We trained the model and evaluated the optimization techniques in terms of convergence speed, accuracy and loss function using three randomly selected publicly available image classification datasets. The overall experimental results obtained show Nadam achieved better performance across the three datasets in comparison to the other optimization techniques, while AdaDelta performed the worst.