Adam and RMSProp are two of the most influential adaptive stochastic algorithms for training deep neural networks, which have been pointed out to be divergen