This paper first answers the question "why do the two most powerful techniques Dropout and Batch Normalization (BN) often lead to a worse performance when