Achieving good speed and accuracy trade-off on a target platform is very important in deploying deep neural networks in real world scenarios. However, most e