

In this paper we propose the macroblock scaling (MBS) algorithm, which can be applied to various CNN architectures to reduce their model size. MBS adaptively