

# Supplementary Material for Block-wisely Supervised Neural Architecture Search with Knowledge Distillation

Changlin Li<sup>2\*</sup>, Jiefeng Peng<sup>1\*</sup>, Liuchun Yuan<sup>1,3</sup>, Guangrun Wang<sup>1,3†</sup>, Xiaodan Liang<sup>1,3</sup>,  
Liang Lin<sup>1,3</sup>, Xiaojun Chang<sup>2</sup>

<sup>1</sup>DarkMatter AI Research <sup>2</sup>Monash University <sup>3</sup>Sun Yat-sen University

{changlin.li, xiaojun.chang}@monash.edu, {jiefengpeng, ylc0003, xdliang328}@gmail.cn,  
wanggrun@mail2.sysu.edu.cn, linliang@ieee.org

## A. Appendix

### A.1. Model Architectures

Fig 1 shows the details of our searched architectures. ‘MB  $x y \times y$ ’ stands for an Inverted bottleneck convolution module with expand rate  $x$  and kernel size  $y$ . Blue blocks represent blocks with expand rate 6, whereas green ones represent blocks with expand rate 3. Orange blocks are fixed during architecture search.

Searched without constraint, DNA-d selects relatively expensive operations with large kernel size and high expansion rate. DNA-c, under the constraint of parameter number, tends to lower model size by reduce expansion rate and layer number at the tail end. Under the constraint of computational complexity, DNA-b and DNA-a tend to choose operations with fewer channels and lower expansion rate evenly in each block.

---

\*Changlin Li and Jiefeng Peng contribute equally and share first-authorship. This work was done when Changlin Li worked as an intern at DarkMatter AI.

†Corresponding Author is Guangrun Wang.

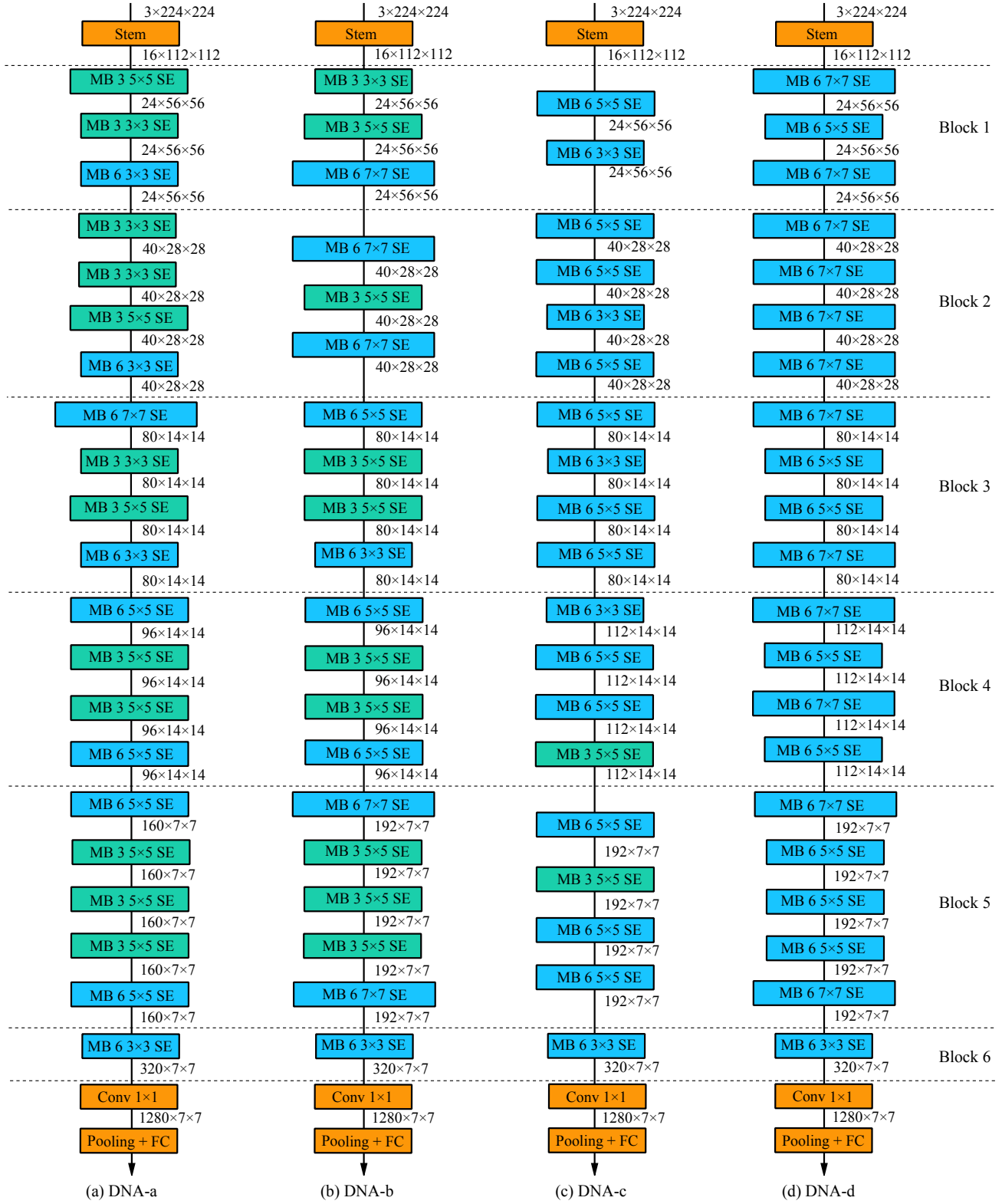


Figure 1. Architectures of DNA-a,b,c,d. ‘MB  $x y \times y$ ’ stands for an Inverted bottleneck convolution module with expand rate  $x$  and kernel size  $y$ . Blue blocks represent blocks with expand rate 6, whereas green ones represent blocks with expand rate 3. Orange blocks are fixed during architecture search.