

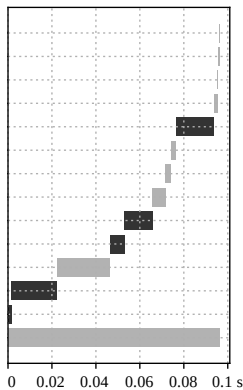
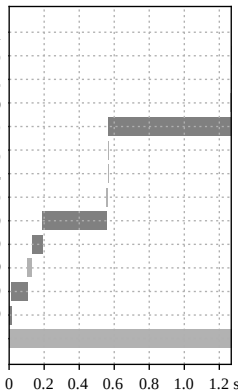
Model B (floating-point)

Tensor operation

CPU

CPU + TP
(hybrid custom
floating-point)

SOFTMAX
FULLY_CONNECTED
RESHAPE
MAX_POOL_2D
(5B) CONV_2D
ADD
MUL
MAX_POOL_2D
(4B) CONV_2D
(3B) D_CONV_2D
MAX_POOL_2D
(2B) CONV_2D
(1B) D_CONV_2D
Interpreter



Time (s)