

人工智慧晶片設計與應用

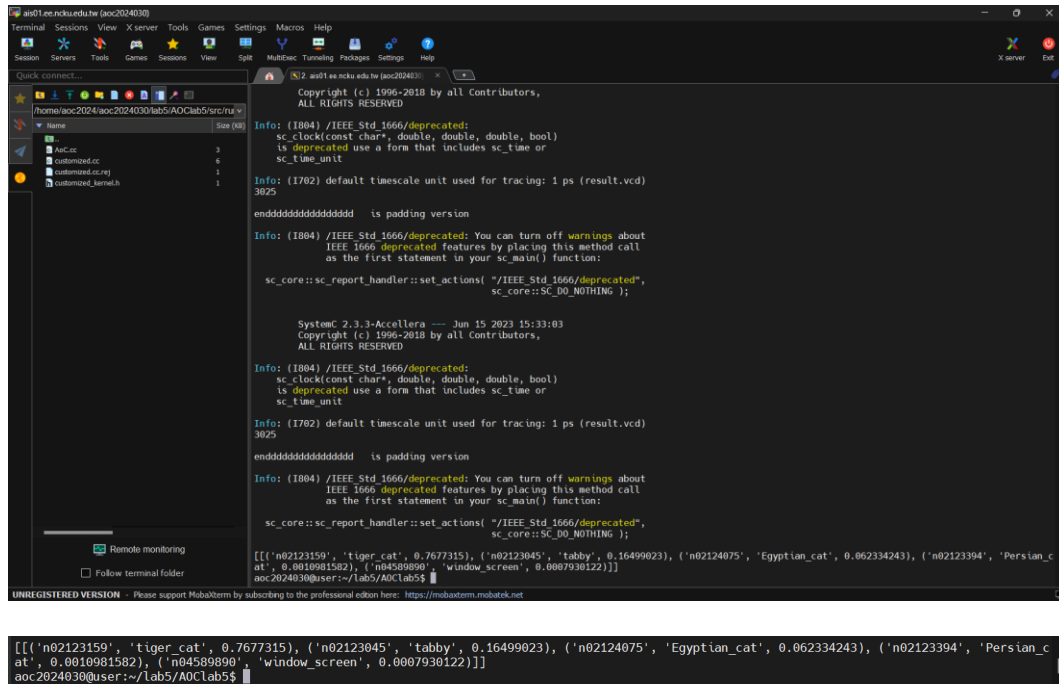
AI-ON-CHIP FOR MACHINE LEARNING AND INFERENCE

Lab 5

E14096724

鄭喆嚴

1. Pass: terminal result after run “python3 gallery/tutorial/customized.py” as same as the value below



```
Copyright (c) 1996-2018 by all Contributors,
ALL RIGHTS RESERVED

Info: (I804) /IEEE_Std_1666/deprecated:
sc_clock(const char*, double, double, bool)
is deprecated use a form that includes sc_time or
sc_time_unit

Info: (I702) default timescale unit used for tracing: 1 ps (result.vcd)
3025

enddddddddddddd is padding version

Info: (I804) /IEEE_Std_1666/deprecated: You can turn off warnings about
IEEE 1666 deprecated features by placing this method call
as the first statement in your sc_main() function:

sc_core::sc_report_handler::set_actions( "/IEEE_Std_1666/deprecated",
sc_core::SC_DO_NOTHING );

SystemC 2.3.3-Accellera --- Jun 15 2023 15:33:03
Copyright (c) 1996-2018 by all Contributors,
ALL RIGHTS RESERVED

Info: (I804) /IEEE_Std_1666/deprecated:
sc_clock(const char*, double, double, bool)
is deprecated use a form that includes sc_time or
sc_time_unit

Info: (I702) default timescale unit used for tracing: 1 ps (result.vcd)
3025

enddddddddddddd is padding version

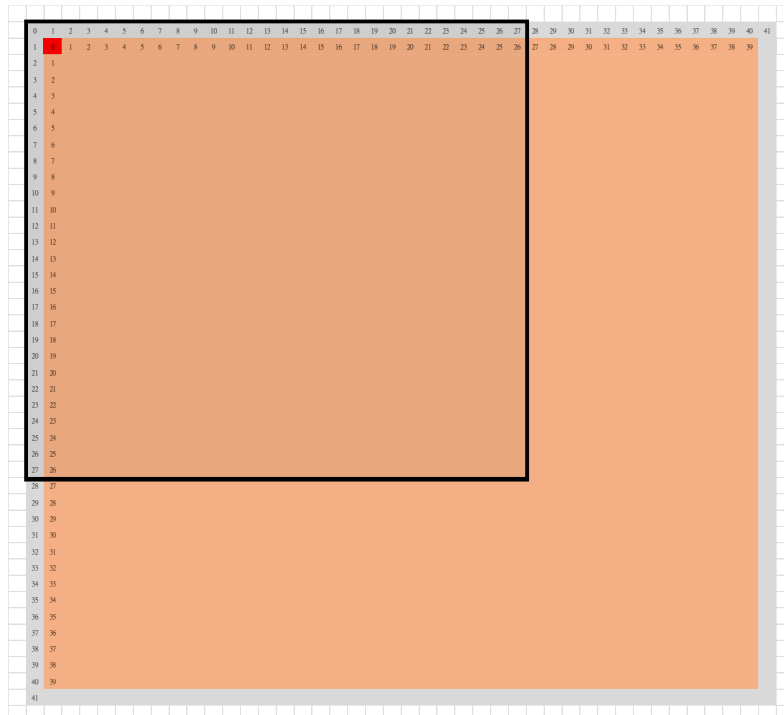
Info: (I804) /IEEE_Std_1666/deprecated: You can turn off warnings about
IEEE 1666 deprecated features by placing this method call
as the first statement in your sc_main() function:

sc_core::sc_report_handler::set_actions( "/IEEE_Std_1666/deprecated",
sc_core::SC_DO_NOTHING );

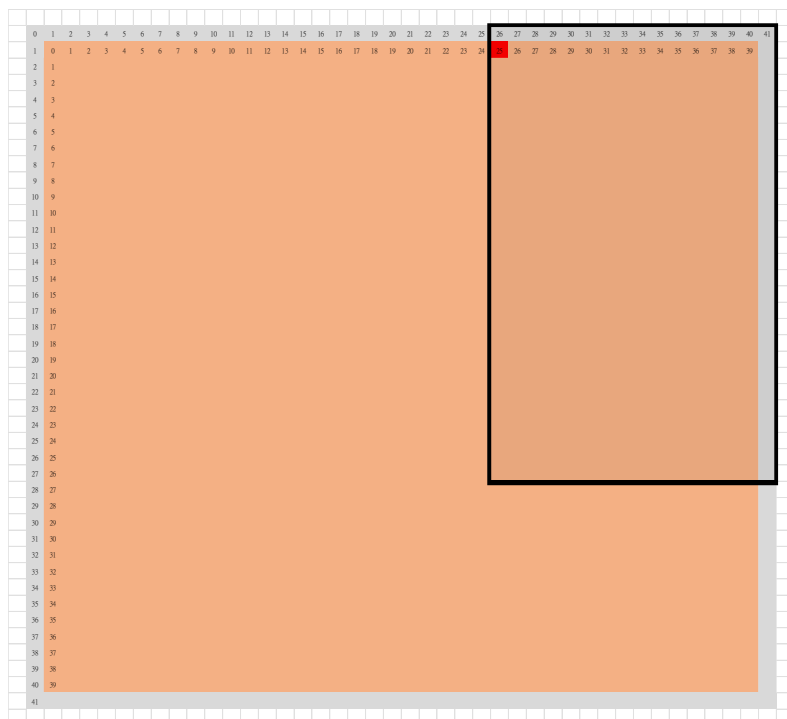
[[('n02123159', 'tiger_cat', 0.7677315), ('n02123045', 'tabby', 0.16499023), ('n02124075', 'Egyptian_cat', 0.062334243), ('n02123394', 'Persian_cat', 0.0010981582), ('n04589890', 'window_screen', 0.0007930122)]]
aoc2024030@user:~/lab5/AOClab5$
```

2. Explain how you tile feature map?

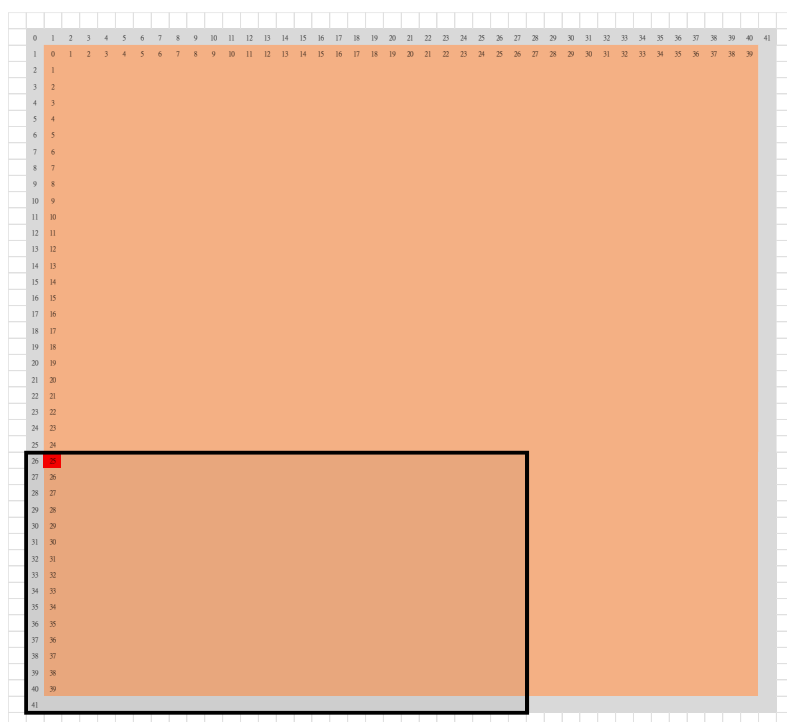
我的 tile size 是 28，假設 padding 過後的 feature map 的寬和高都是 42，先從位置 0 開始 tile，接著往 feature map 的寬進行下一次的 tile。



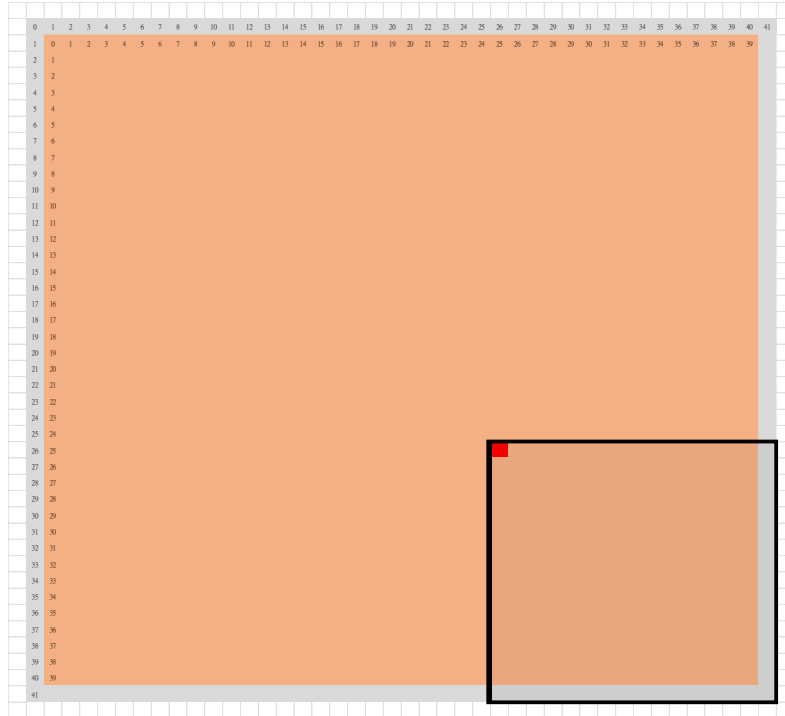
需要和上一次的 tile 重複兩排，因為 filter 是 3×3 ，所以下次 tile 時還是會用到上次 tile 的最後兩排，如果 feature map 剩餘的寬度小於等於 tile size，就把 tile 的寬度調整成剩餘的寬度，並且回到一開始 tile 的位置往 feature map 的高進行下一次的 tile。



需要和上一次的 tile 重複兩列，因為 filter 是 3×3 ，所以下次 tile 時還是會用到上次 tile 的最後兩列，如果 feature map 剩餘的高度小於等於 tile size，就把 tile 的高度調整成剩餘的高度，並且往 feature map 的寬進行下一次的 tile。



需要和上一次的 tile 重複兩排，因為 filter 是 3*3，所以下次 tile 時還是會用到上次 tile 的最後兩排，如果 feature map 剩餘的寬度和高度都小於等於 tile size，就把 tile 的寬度和高度都調整成剩餘的寬度和高度，在完成這次 tile 的運算後還下一張 feature map。



3. Explain how you assign DRAM address?

input feature map DRAM address:

將 ifmap 的 address 起始點加上 channel 的 shift，就是當前所在 ifmap 的 channel 乘上 ifmap 的寬乘高，再加上當前在 DRAM 中的高乘 ifmap 的寬，再加上當前在 DRAM 中的寬。當前在 DRAM 中的高要判斷是否有 ifmap 上方的 padding 並且判斷當前 tile 的高度位置使否在 0 來得到當前 tile 起始的高度位置對應到 DRAM 中的高；當前在 DRAM 中的寬要判斷是否有 ifmap 左邊的 padding 並且判斷當前 tile 的寬度位置使否在 0 來得到當前 tile 起始的寬度位置對應到 DRAM 中的寬。

$$\begin{aligned} \text{dram_addr} = & \text{DRAM_INPUT_START_ADDR} + \\ & \text{ifmap_W} * \text{ifmap_H} * k + \\ & \text{ifmap_i} + \text{ifmap_j} * \text{ifmap_W}; \end{aligned}$$

output feature map DRAM address:

將 ofmap 的 address 起始點加上 ofmap 張數的 shift，就是當前所在 ofmap 的張數乘上 ofmap 的寬乘高，再加上當前 tile 起始的高度位置乘 ofmap 的寬，再加上 tile 起始的寬度。

$$\begin{aligned} \text{dram_addr} = & \text{DRAM_OUTPUT_START_ADDR} + \\ & (\text{padded_w} - \text{filter_W} + 1) * (\text{padded_h} - \text{filter_H} + 1) * \text{out_c} + \\ & i + j * (\text{padded_w} - \text{filter_W} + 1); \end{aligned}$$

4. Lesson Learned

這次的 Lab 讓我學習到如何進行資料的 tiling，如何調整 tiling 的 window 才能正確的拿取資料，像是要設計在 tiling 的 window 超出 feature map 的時候要如何調整拿取資料的寬度，還要設計 padding 後的 feature map 的 address 要如何調整，我覺得在設計 feature map 和 output feature map 的地址時是最困難的，這考驗是否了解整個 convolution 的運作，計算 address 時還要考慮到一開始 feature map 的 padding 後調整，我覺得這次的 Lab 有讓我更加了解有關 AI compiler 如何進行 feature map 的 tiling。