

1-D Winograd: $F(2,3)$

1	2	3	4
---	---	---	---

input feature map

0	1	0
---	---	---

kernel

2	3
---	---

output feature map

Winograd 计算式: $Y = A^T [(Gg) \odot (B^T d)]$

2-D Winograd: $F(2 \times 2, 3 \times 3)$

0	1	1	0
1	0	1	0
0	0	0	1
0	1	0	0

input feature map

0	0	0
0	0	1
0	0	0

kernel

1	0
0	1

output feature map

(from ppt: "a minimal 1-D algorithm is nested with itself to obtain a 2-D algorithm")

Winograd 计算式: $Y = A^T [GG^T] \odot [B^T d B] A$

现在 input map $H=W=56$, kernel 仍是 3×3 , padding = 0. stride = 1

那么 output feature map 应该是 $\frac{56-3+0}{1} + 1 = 54 \times 54$

(假设 input channel = output channel = batch size = 1)

所以我们需要 $F(54 \times 54, 3 \times 3)$, 而这不能用 $F(2 \times 2, 3 \times 3)$ 的方法做的, 所以 PPT 上

用了分块的方法 (其实你也可借助工具生成 $F(54 \times 54, 3 \times 3)$ 的 A, B, G 矩阵直接做?)

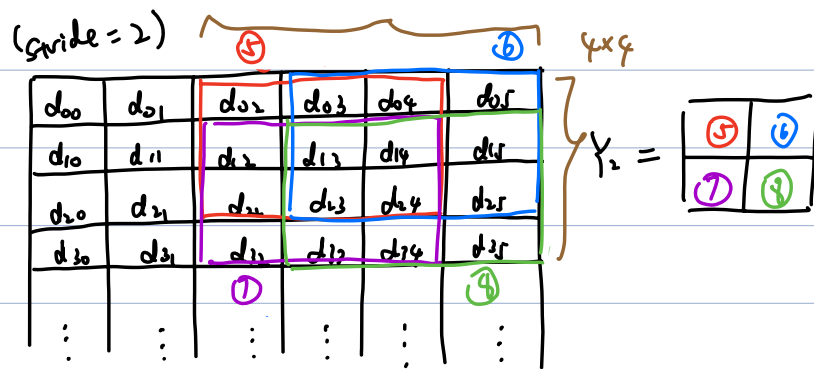
(附: 分块卷积 3×3 kernel)

d ₀₀	d ₀₁	d ₀₂	d ₀₃	d ₀₄	d ₀₅
d ₁₀	d ₁₁	d ₁₂	d ₁₃	d ₁₄	d ₁₅
d ₂₀	d ₂₁	d ₂₂	d ₂₃	d ₂₄	d ₂₅
d ₃₀	d ₃₁	d ₃₂	d ₃₃	d ₃₄	d ₃₅
⋮	⋮	⋮	⋮	⋮	⋮

input map

把 input map 分成 4×4 的小块

①	②
③	④



native 卷积结果就等于各块卷积结果的合并. 即 $(Y_1 \parallel Y_2)$

最后, 再考虑 input channel, output channel 和 batch 的问题. 实际上就是在加几层循环。