Difference between OS and Inception

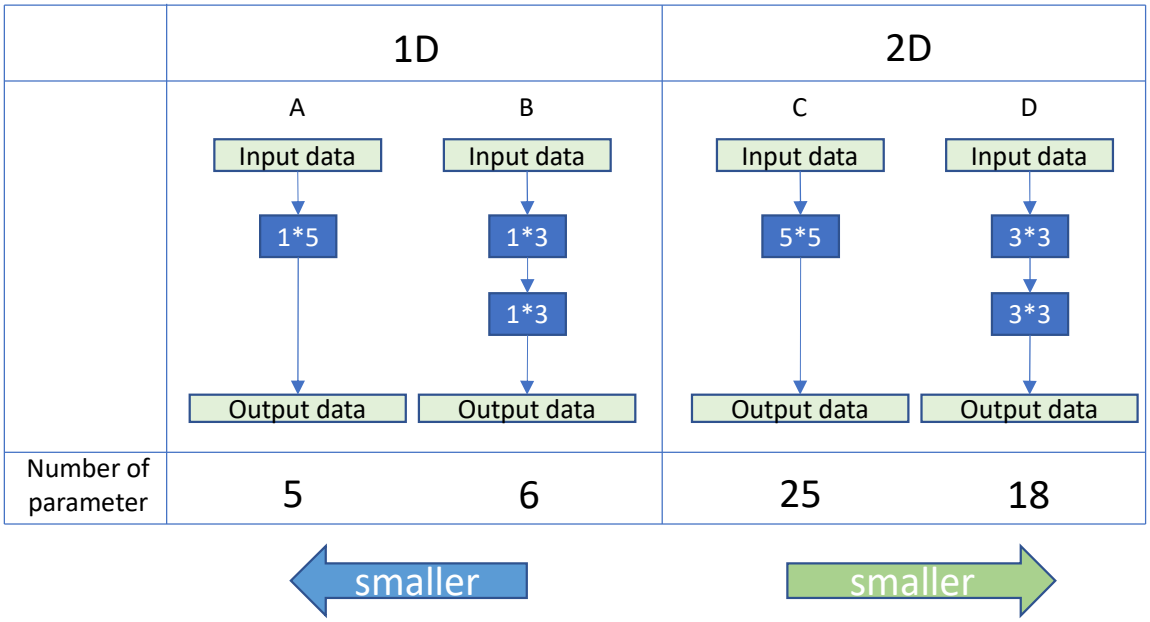
You might think OS design is a type of Inception.
It is true. Omin-scale is a type of multi-scale.
However, there are something behind it.

To summarise, OS is a **shallower** design (it should be)
OS is built by a **unique** model size reduction method.

1D-CNN tends to be shallower

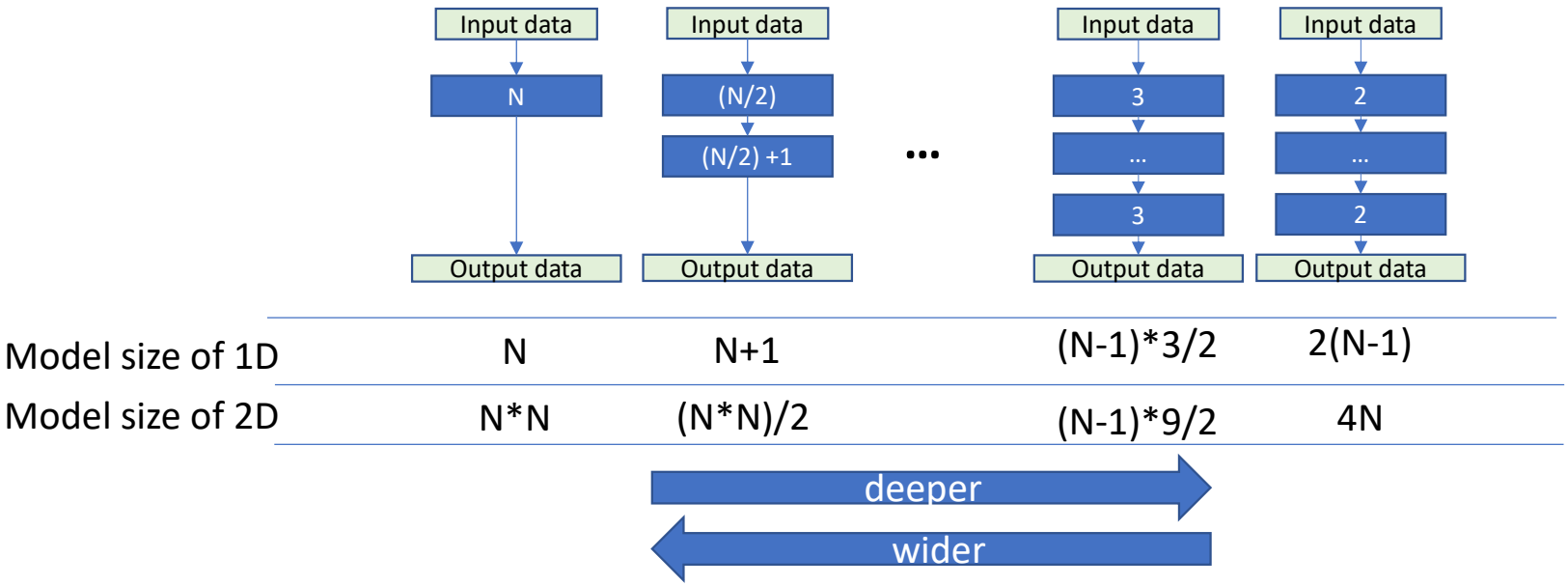
Start from a simple case

- In 1D-CNN kernel factorization cannot reduce mode size as 2D-CNN. It will increase the model size.



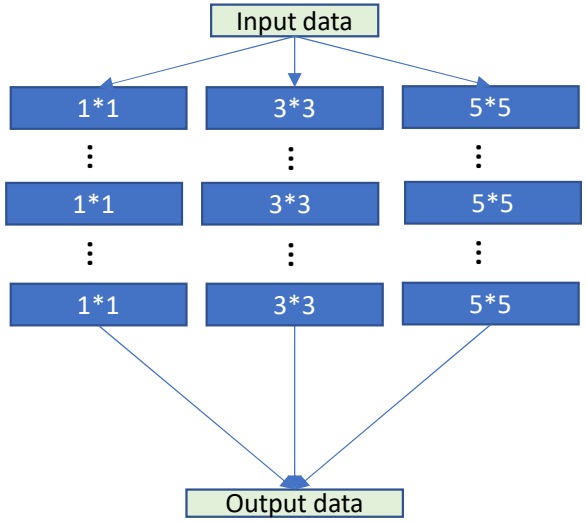
Supposing the ideal size is N

- For 1D-CNN, going deeper will increase the model size
- For 2D-CNN, going deeper will reduce the model size



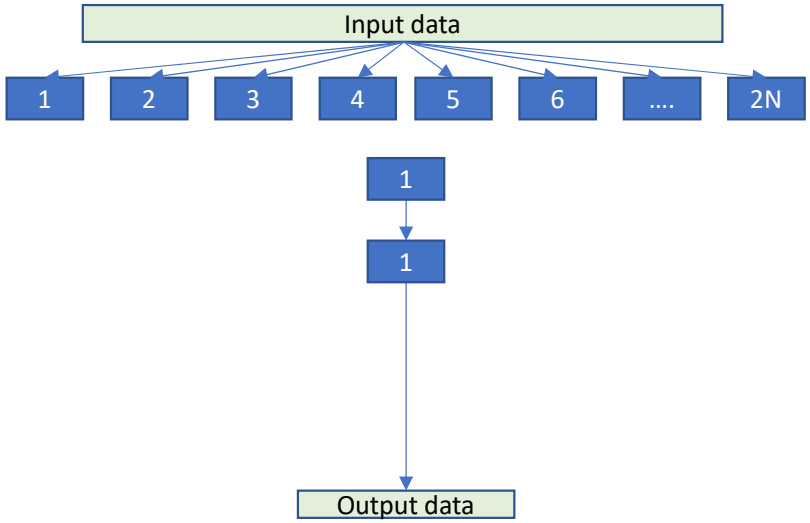
For smaller model size

2D-CNN should be
Stack of many layers of small kernels to
build RF of different sizes



Inception

1D-CNN should be
Stack of few layers of various kernels to
have RF of different size
Following layer is used for add non-linear



All-RF design

The First difference of inception and OS

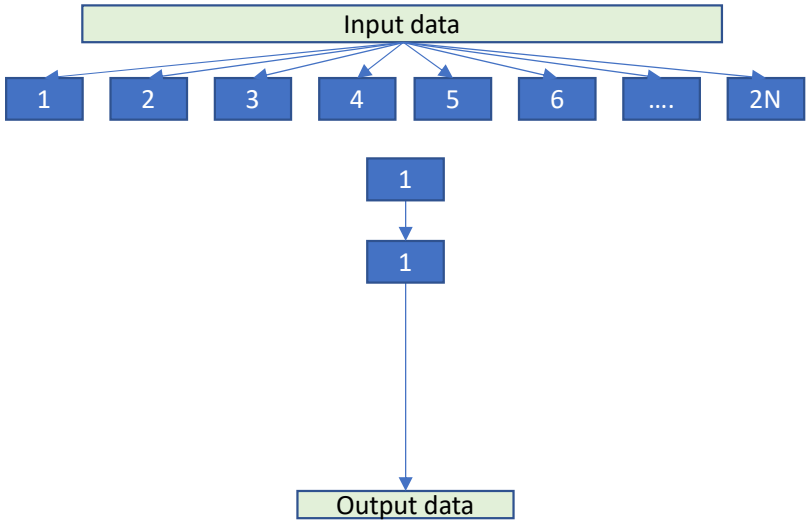
- For 2D-CNN, the smaller model size solution to build RF of different size is:
  - stacking many thinner layer to build RF of difference sizes.
    - (stacking will make it **deeper**)

- For 1D-CNN, the smaller model size solution to build RF of different size is:
  - Just using RF of difference sizes.
    - (just using will make it **shallower**)

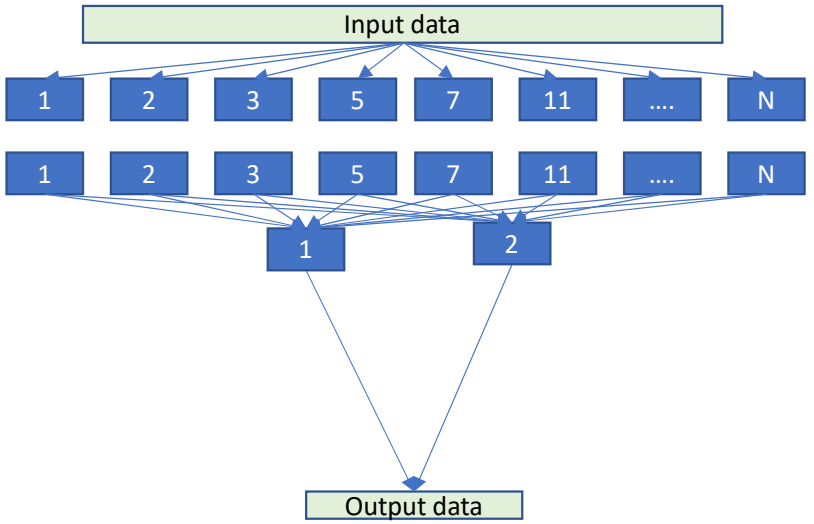
OS is built by a unique model size reduction method.

Can we make the model size more smaller?

All-RF



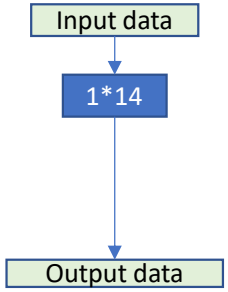
OS-CNN



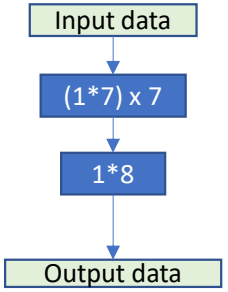
In order to learn same information as the All-RF one
The OS-CNN need to have more channel number for each kernel. Therefore, we cannot merely calculate mode size by sum the kernel size.

What's the channel number should be?

- As the subsection "**No representation ability lose**"[1][2] says
  - If we want to do kernel factorization we need to increase the channel number of first layer
  - And the channel number is min(A,B). Where a and b are kernel sizes. The example is below:

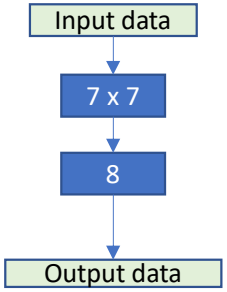


A



B

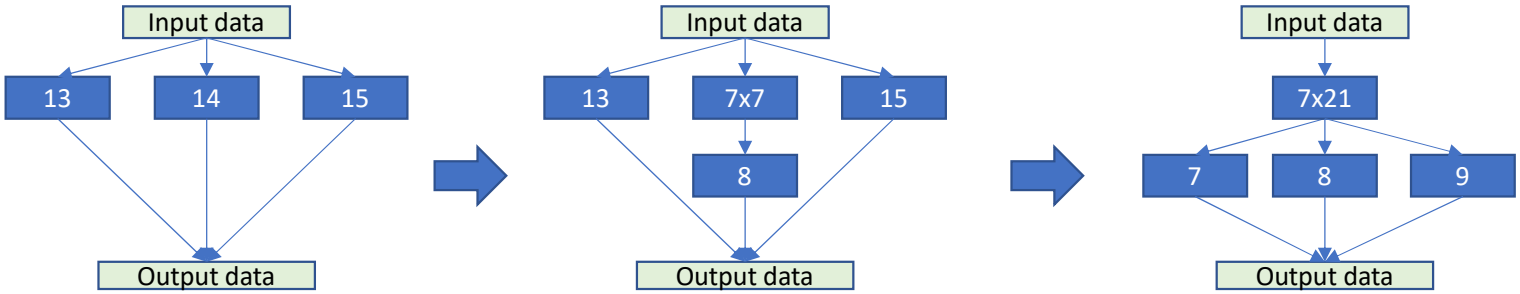
From A to B the channel requirement for 1*7 kernel is 7
Let's write it in and remove 1* for the following discussion is all based on 1D-CNN. Then we should write B as C



C

[1] https://github.com/Wensi-Tang/OS-CNN/blob/master/Code_example_of_theoretical_proof/4_3_Check_No_representation_ability_lose.ipynb
[2] https://github.com/Wensi-Tang/OS-CNN/blob/master/Appendix/Proof%20of%20No%20representation%20ability%20lose.pdf

Simple factorization will still increase model size

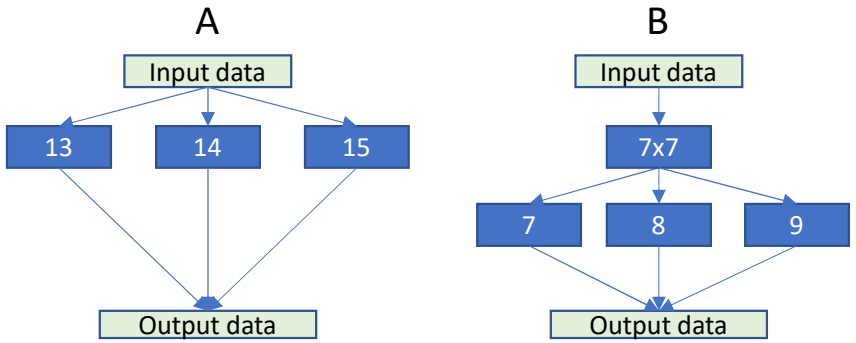


Number of parameter	42	13+8+15+7*7 = 85	7+8+9+7*21 = 171

However

- The objective that we want kernels of all size is:
  - we want to find the proper kernel.
  - Therefore, we don't really need 21 channels.

- For example, if we just want 1 kernel. Then, the model should selected from {13, 14, 15} during training time, and model B should be selected from {7, 8, 9}. And for network B, only 7 channels would be enough!



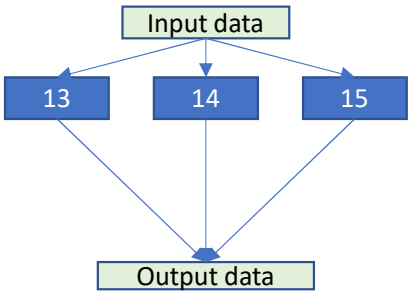
Number of parameter	42	7+8+9+7*7 = 73

It seems A is still of smaller number of parameter than B, But…..

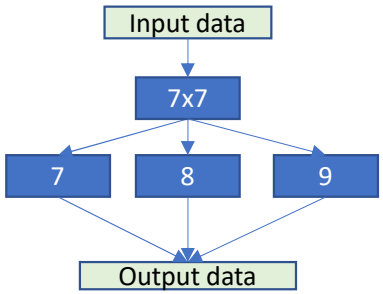
If we don’t want to select from {13, 14, 15}, if we want to select from 8 to 20

This is of smaller number of parameter!

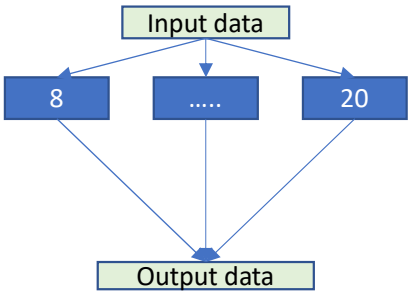
A



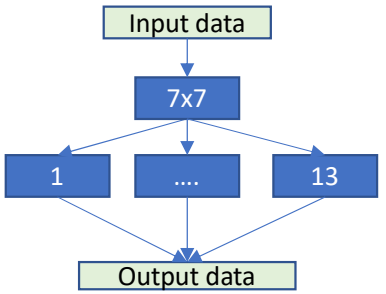
B



A’



B’



	A	B	A’	B’
Number of parameter	42	7+8+9+7*7 = 73	8+9+,…,+20 = 182	(1+2,…,+13)+49= 140

The Second difference of OS and inception

- This kind of model size reduction does not even exist in 2D-CNN.