

Why Deep Learning?

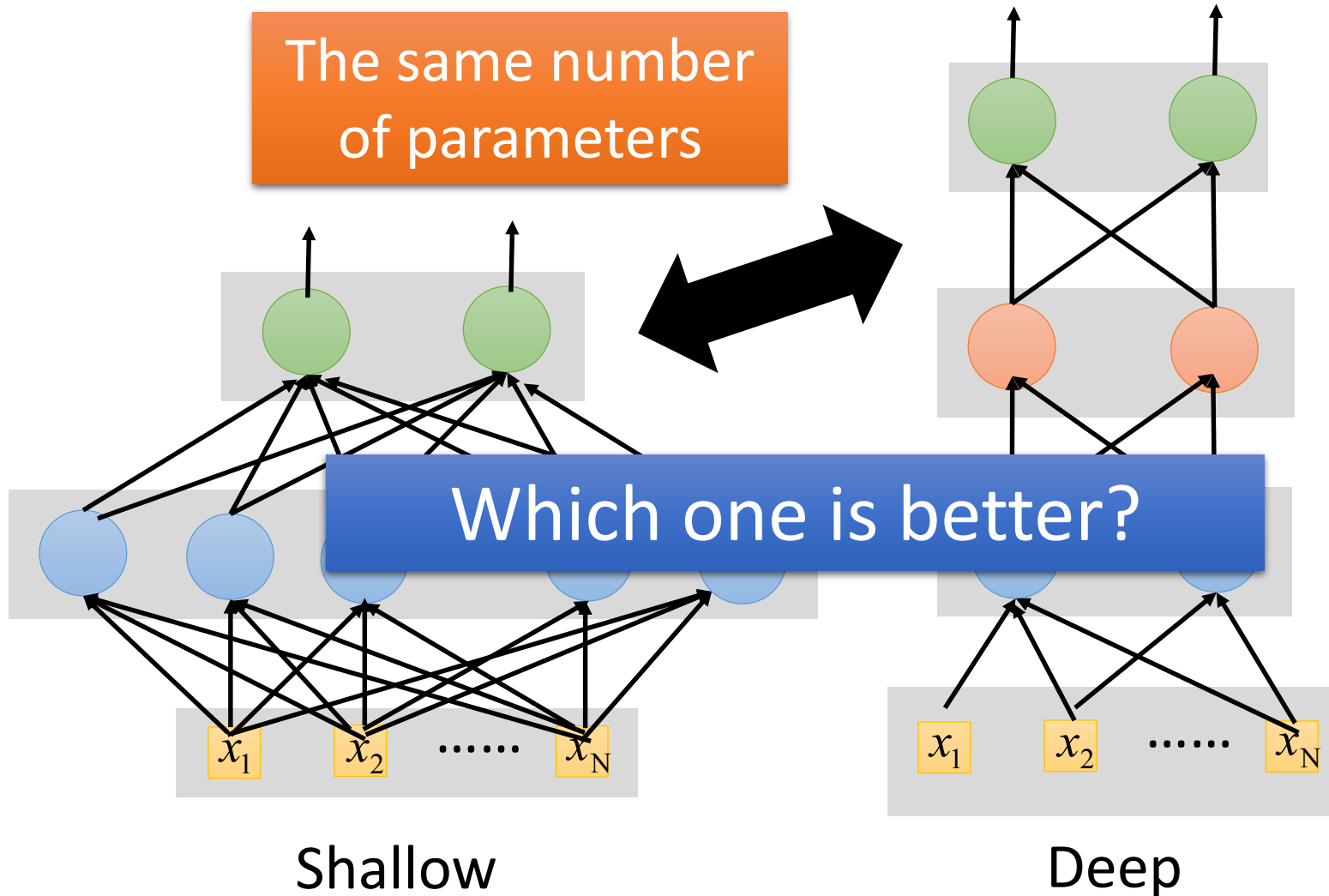
Deeper is Better?

Layer X Size	Word Error Rate (%)
1 X 2k	24.2
2 X 2k	20.4
3 X 2k	18.4
4 X 2k	17.8
5 X 2k	17.2
7 X 2k	17.1

Not surprised, more parameters, better performance

Seide, Frank, Gang Li, and Dong Yu. "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks." *Interspeech*. 2011.

Fat + Short v.s. Thin + Tall



Fat + Short v.s. Thin + Tall

發現：在參數量相同的情況下，Deep 比 Fat 表現得更

Layer X Size	Word Error Rate (%)	Layer X Size	Word Error Rate (%)
1 X 2k	24.2		
2 X 2k	20.4		
3 X 2k	18.4		
4 X 2k	17.8		
5 X 2k	17.2	1 X 3772	22.5
7 X 2k	17.1	1 X 4634	22.6
		1 X 16k	22.1

Why?

↔

↔

Seide, Frank, Gang Li, and Dong Yu. "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks." *Interspeech*. 2011.

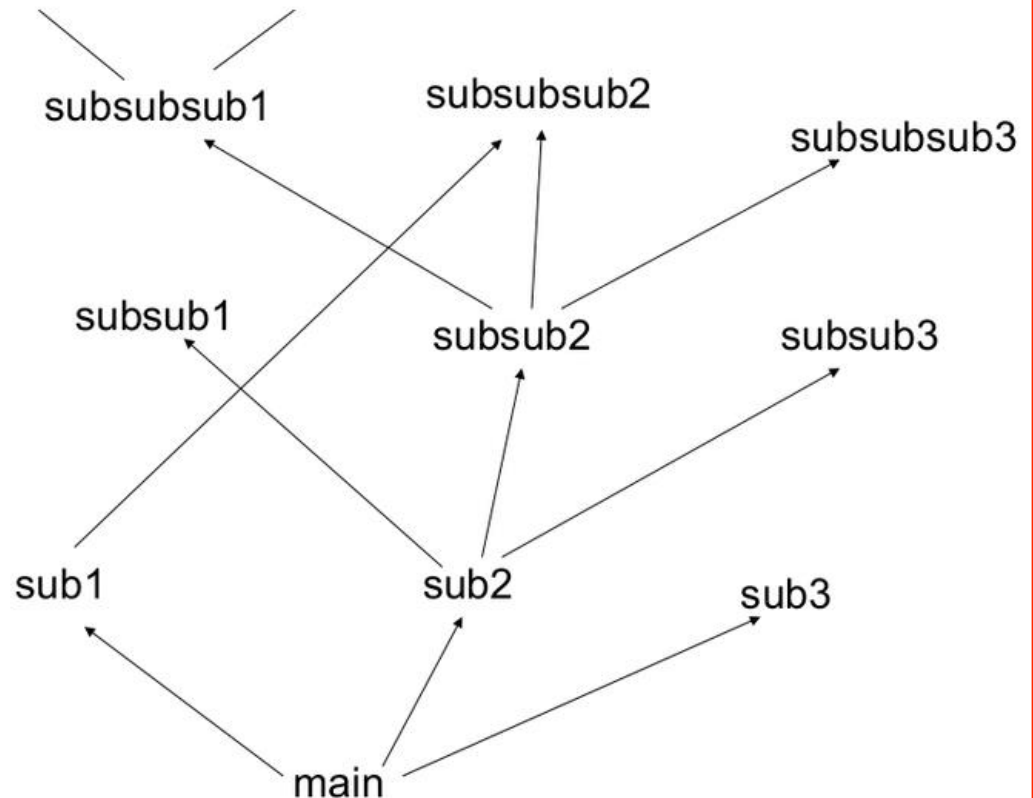
Modularization

關鍵：Deep 象徵「模組化」的概

模組化概念：在寫程式時，並不會將所有的東西寫在 main function 中，而是會將一些常用的功能包裝成 function，以利重複使用

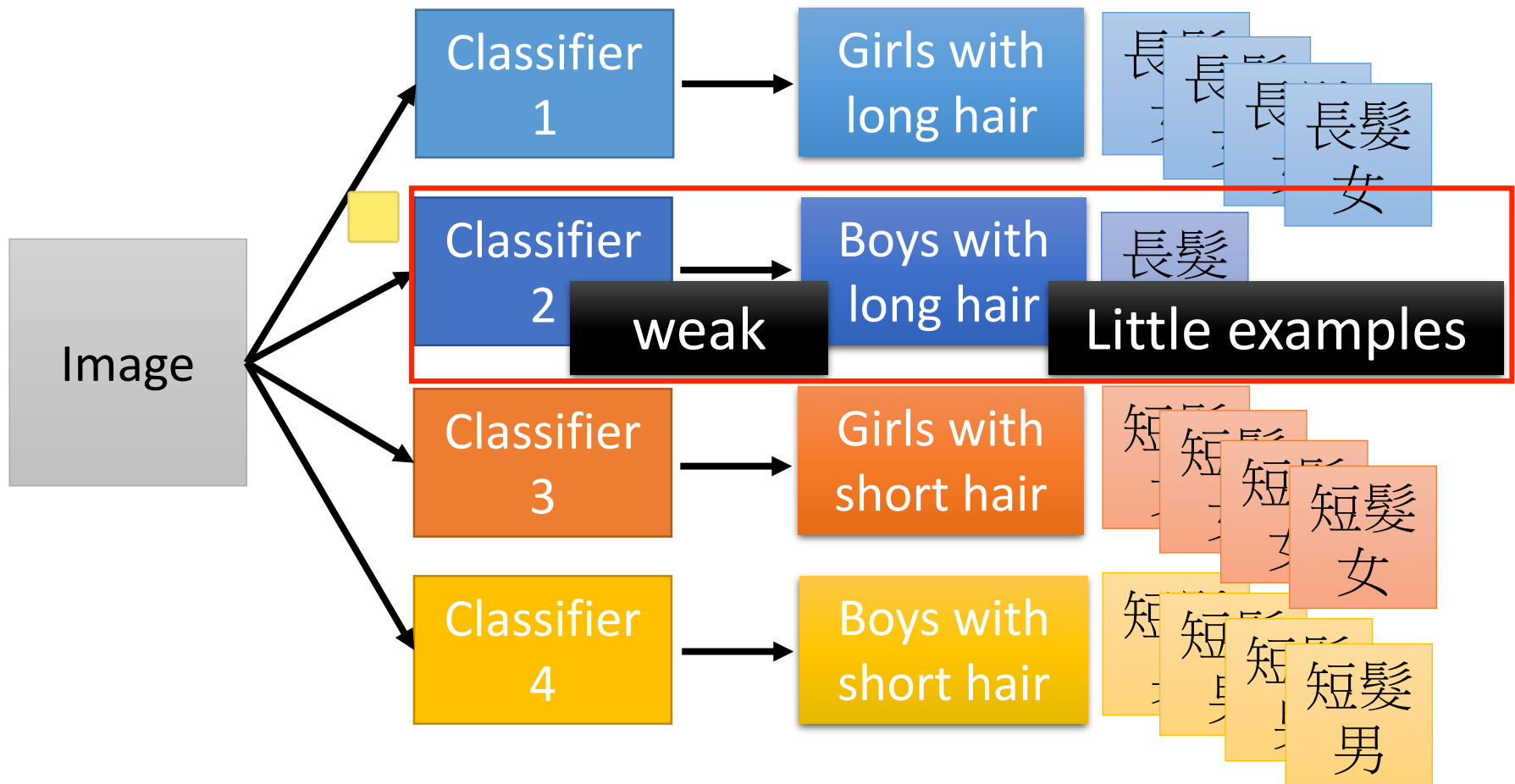
- Deep → Modularization

Don't put everything in your main function.



Modularization

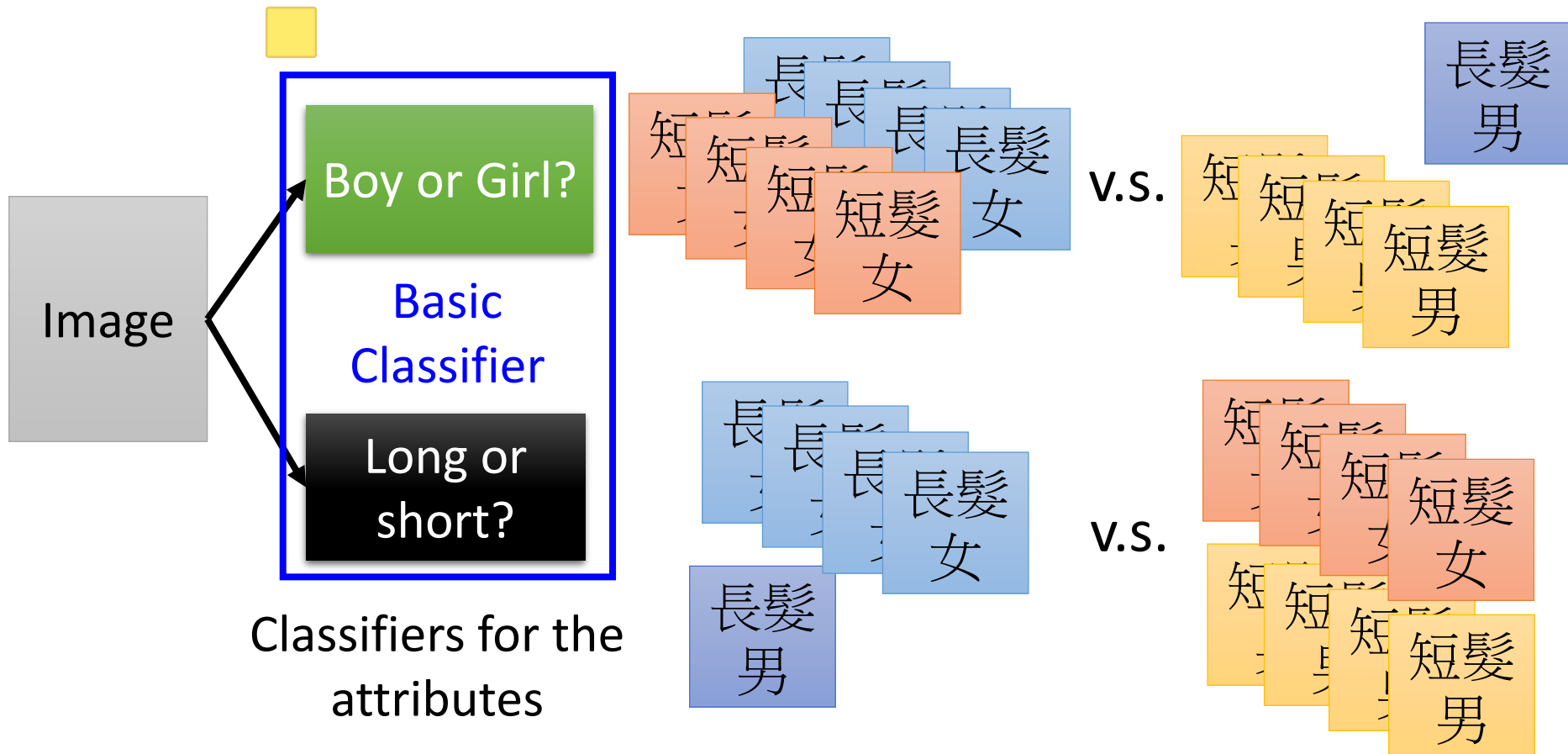
- Deep → Modularization



Modularization

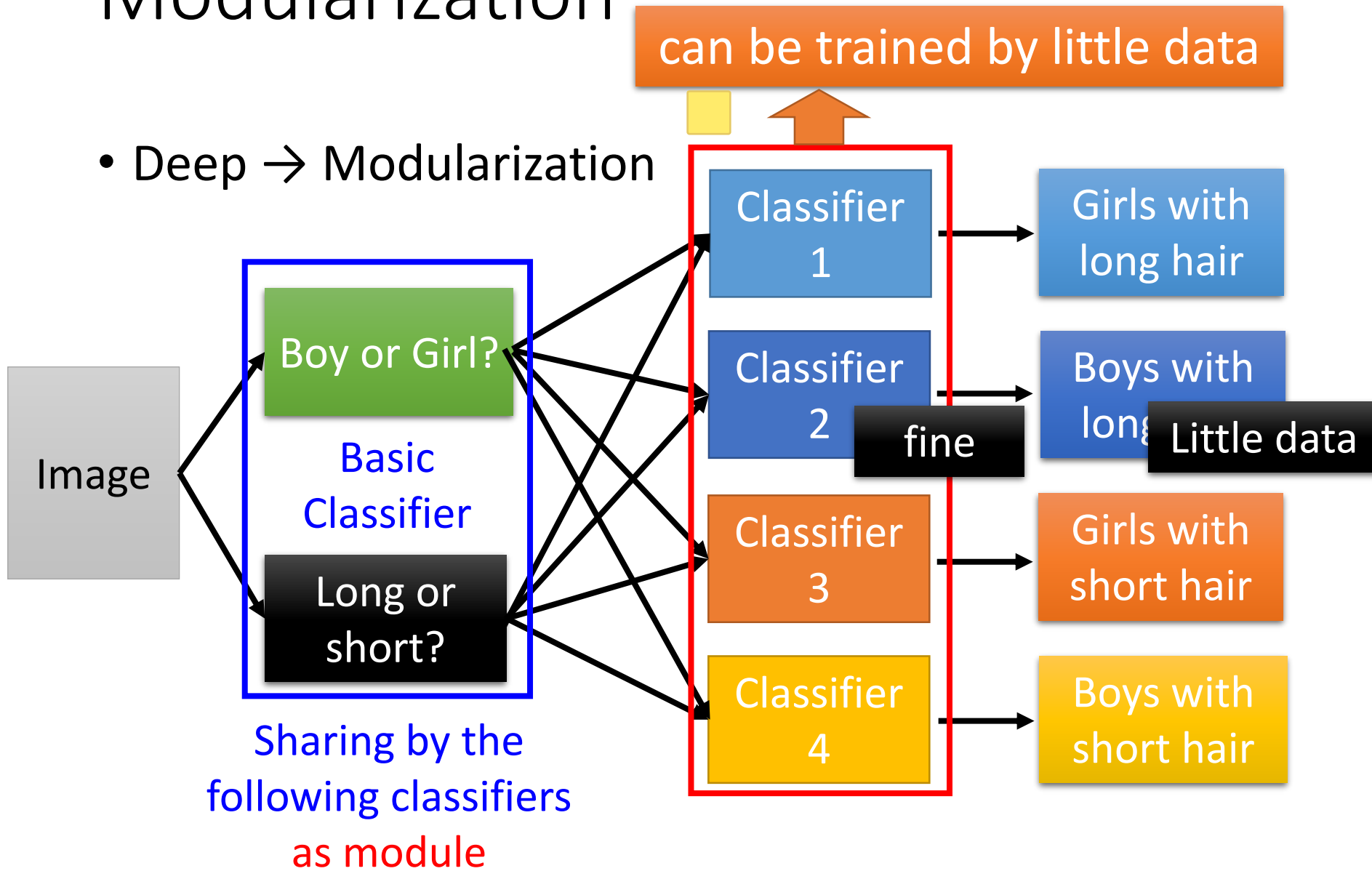
Each basic classifier can have sufficient training examples.

- Deep → Modularization



Modularization

- Deep → Modularization

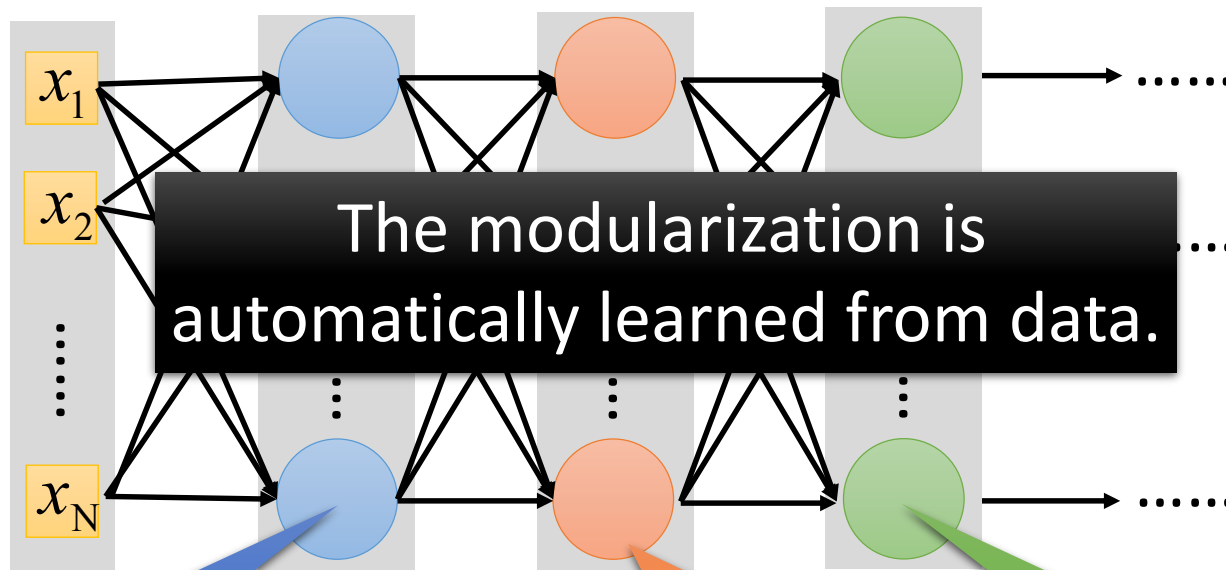


Modularization

所以 Deep Learning 不單純只是利用一個很複雜的 Model (參數很多)

然後利用很多 Training Data 硬 Train 下去! Deep Learning 實際用

- Deep → Modularization → Less training data?



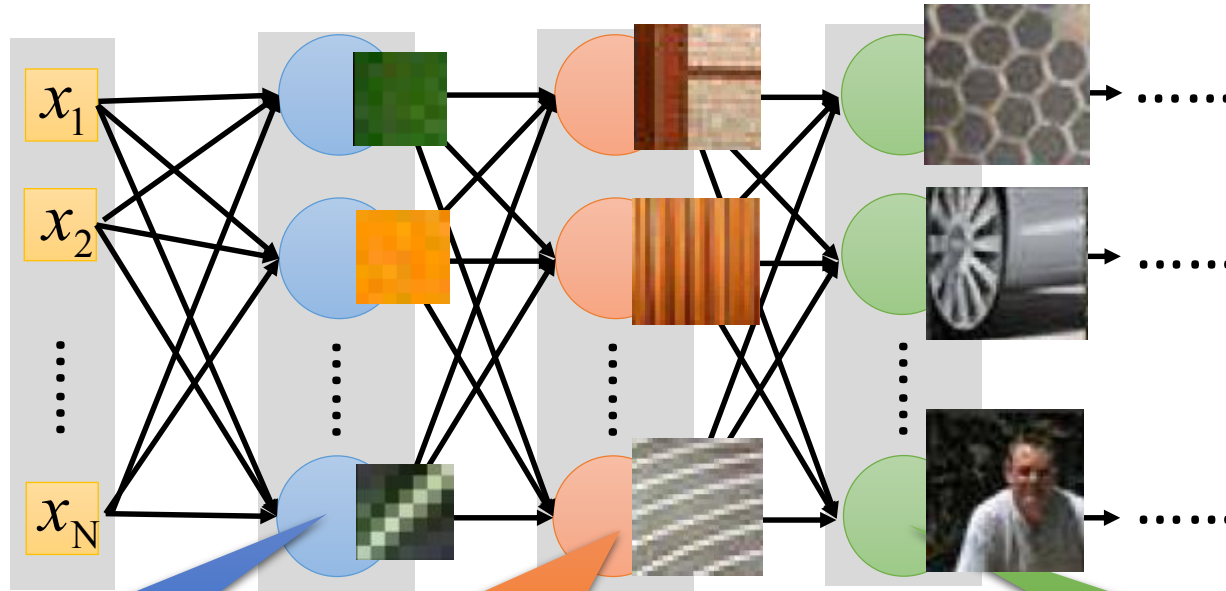
The most basic classifiers

Use 1st layer as module to build classifiers

Use 2nd layer as module

Modularization - Image

- Deep \rightarrow Modularization



The most basic
classifiers

Use 1st layer as module
to build classifiers

Use 2nd layer as
module

Reference: Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision—ECCV 2014* (pp. 818-833)