



深度學習系統與實現

Lab 3-2 – Model Scaling

Dept. of Computer Science and
Information Engineering

National Chiao Tung University



Outline

- Concepts overview
 - Model Scaling
 - EfficientNet
- Lab 3-2 specification
- Grading
- Notices & Hints

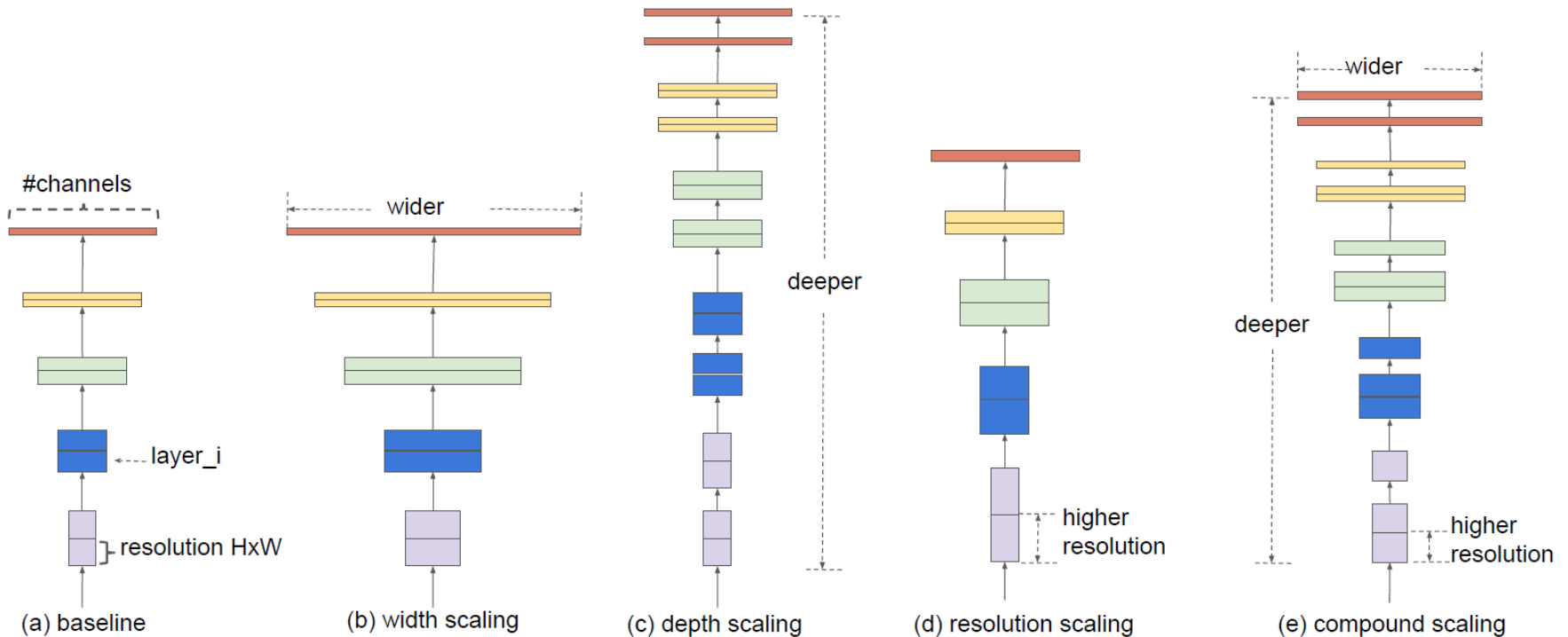


Model Scaling

- Scaling up ConvNets is widely used to achieve better accuracy
 - e.g. ResNet can be scaled up from ResNet-18 to ResNet-200 by using more layers
- The most common ways to scale up ConvNets:
 - width / depth / resolution
- Is there a principled method to scale up ConvNets that can achieve better accuracy and efficiency?



Model Scaling





Compound Model Scaling

□ EfficientNet

- Tan, Mingxing & Le, Quoc. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.

□ Observations

- Scaling up any dimension of network improves accuracy, but the accuracy gain diminishes for bigger models
- Different scaling dimensions are not independent
- It is critical to balance all dimensions of network width, depth, and resolution during scaling



Compound Scaling Method

- ϕ : represents the computational resources
- Uniformly scales network in a principled way:

$$\text{depth: } d = \alpha^\phi$$

$$\text{width: } w = \beta^\phi$$

$$\text{resolution: } r = \gamma^\phi$$

$$\text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \ (\alpha \geq 1, \beta \geq 1, \gamma \geq 1)$$

- constraint $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$ such that for any new ϕ , the total FLOPS will approximately increase by 2^ϕ



Compound Scaling Method

□ Example

- For $\phi = 1$ ($2^\phi = 2^1 = 2$ x computational resource), the experiments of the paper found that the best scaling factors are
 - $\alpha = 1.2$
 - $\beta = 1.1$
 - $\gamma = 1.15$



Specification

Dataset - food11

- Food11 Download link - <https://www.kaggle.com/tohidul/food11>
- Note: in lab3, **full or skewed food11** are both ok





Lab 3-2 Model Scaling

- In lab 3-2, we use Resnet as the backbone

- **[70%] Implement a scalable Resnet**

The scalable Resnet should be able to configure different

- depth (number of layers)
 - Only need to implement integer scaling
 - e.g. The layers of resnet18 = [2, 2, 2, 2], then you can do experiments on [1, 1, 1, 1], [3, 3, 3, 3], etc.
 - You can also try asymmetric settings like [3, 4, 6, 3] or reduce layers (such as modify Resnet to [2, 2, 2], [1, 1, 1], [3, 3, 3], etc.)
- width (number of channels)
- resolution (input size)



Lab 3-2 Model Scaling

□ [30%] Experiments & report

- Compare the different scales of your model (accuracy, execution time, memory usage, etc.)

□ Questions

- Do you get better results by scaling up the model?
 - If not, try to explain & analysis the cause
- Does scaling up only 1 factor/uniformly scaling up 3 factors lead to a different result?
- What should we adjust when a model is scaled up?



Grading

- 3-2 Implementation (70%)
 - 3-2 Experiments & report (30%)
 - Submission: source code + report [e3]
 - zip format (ex: DLSR_lab3-2_{group id}.zip)
 - 5% penalty for the wrong submission format
 - Deadline : 2020/04/27, 23:55 (Mon) [3 weeks]
 - Demo: (TAs will announce on New-E3 later)
- Total
100



Reference

- EfficientNet
 - <https://arxiv.org/abs/1905.11946>