國立高雄科技大學
National Kaohsiung University of Science and Technology

Feature Pyramid Networks (FPN) for Object Detection

Tsung-Yi, Piotr Dollar, Ross Girshick, Kiaming He,
Bharath Hariharan, and Serge Belongie

IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR), 2017

Speaker: Shih-Shinh Huang

October 29, 2021

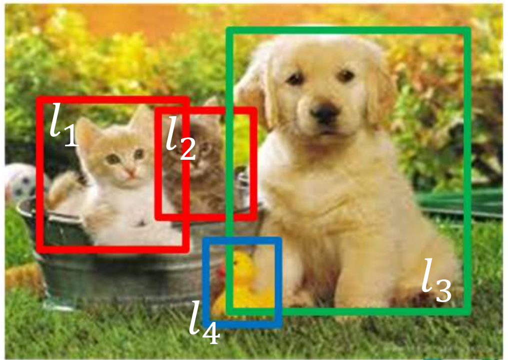
Outline

- Introduction
  - About Object Detection
  - Solution to Multiple Scales
  - Idea of FPN
- Network Architecture
  - Bottom-Up Pathway
  - Lateral Connection
  - Top-Down Pathway
- Region Proposal Network (RPN)
  - About RPN
  - Adopting FPN

Introduction

- About Object Detection
  - Input:
    - I: input image
    - $\{c_1, c_2, \ldots, c_n\}$: object classes to be detected
  - Output:
    - $\{r_1, r_2, \ldots r_m\}$: bounding boxes of m detected objects
    - $\{l_1, l_2, \ldots l_m\}$: class labels of all detected objects

$C = \{\mathbf{cat}, \mathbf{dog}, \mathbf{duck}\}$



$l_1 = \mathbf{cat}$

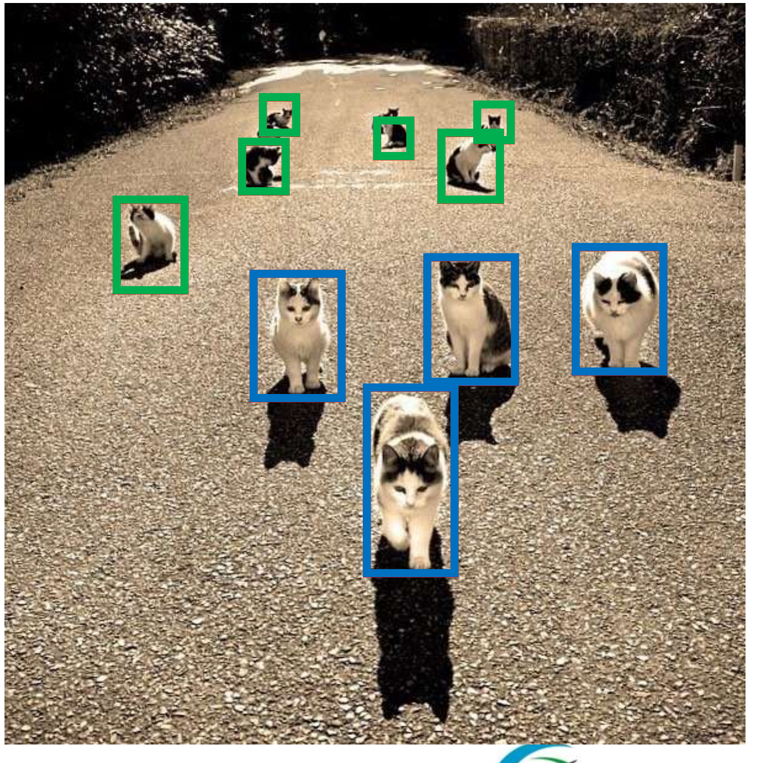
$l_2 = \mathbf{cat}$

$l_3 = \mathbf{dog}$

$l_4 = \mathbf{duck}$

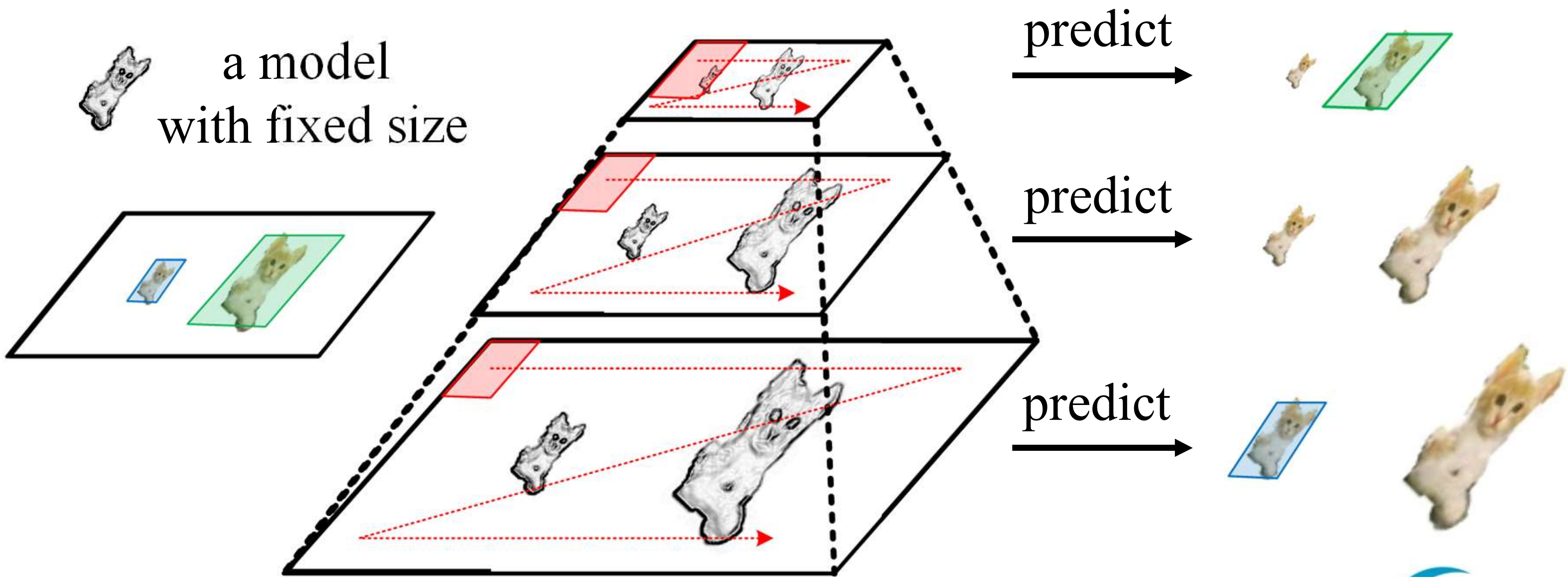
Introduction

- About Object Detection
  - Appearance Variance: object appearances are highly varied.
    - point of view
    - object poses
  - Scale Problem: object sizes are significantly different due to perspective phenomena
    - near → large
    - far → small



Introduction

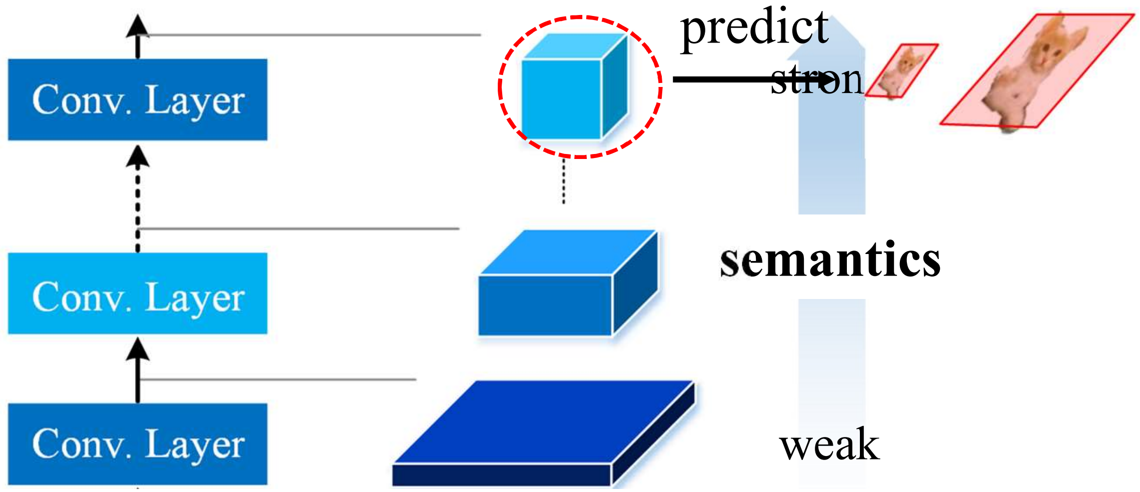
- Solution to Multiple Scales
  - Featurized Image Pyramid: is a basic approach for addressing scale problem



drawback: require dense scale sampling to achieve good results

Introduction

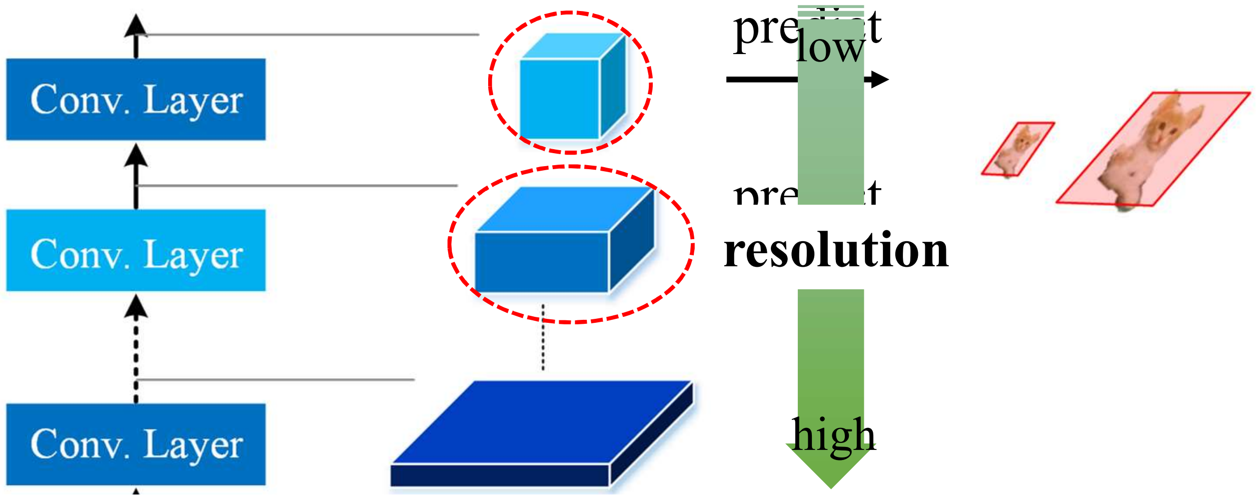
- Solution to Multiple Scales
  - Single Deep Map: only use __a__ feature map from the last Conv. layer for prediction.



pyramids are still required to get more accurate results.

Introduction

- Solution to Multiple Scales
  - Deep Feature Hierarchy: use the feature maps with enough semantics from the high-up layers.

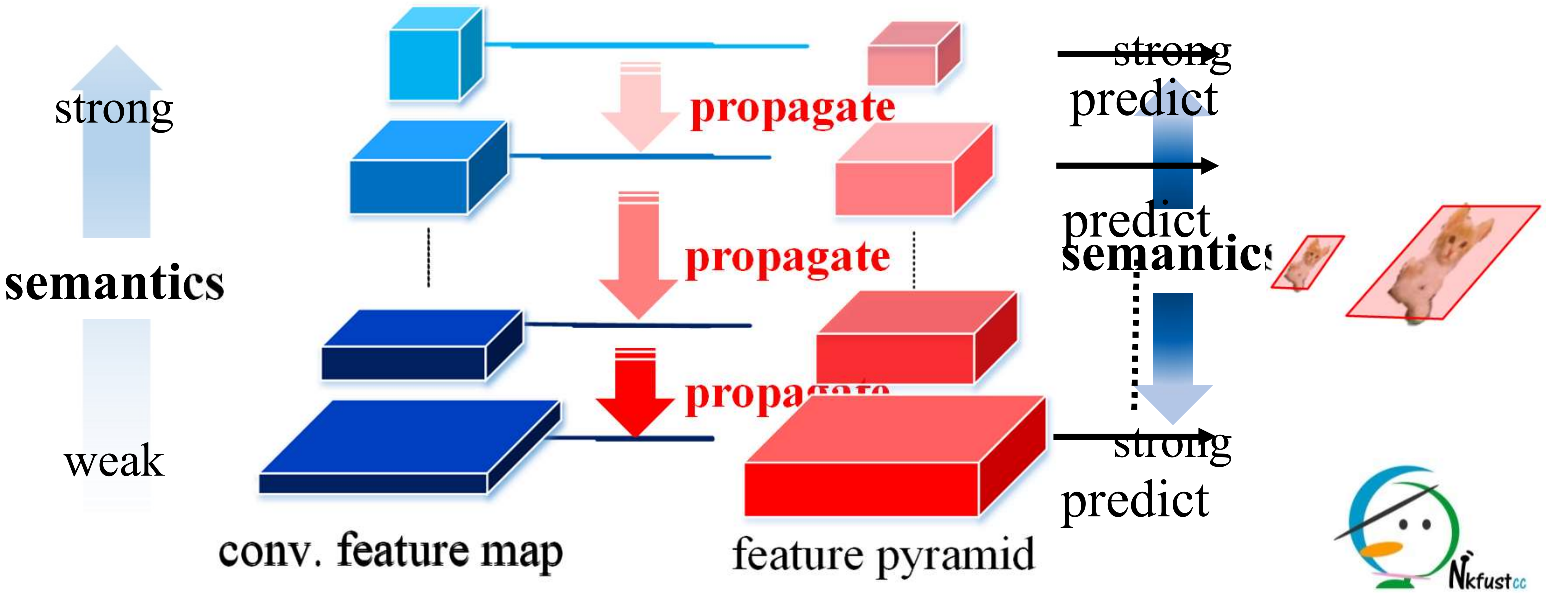


drawback: miss to use high-resolution maps that are important for detecting small objects

Introduction

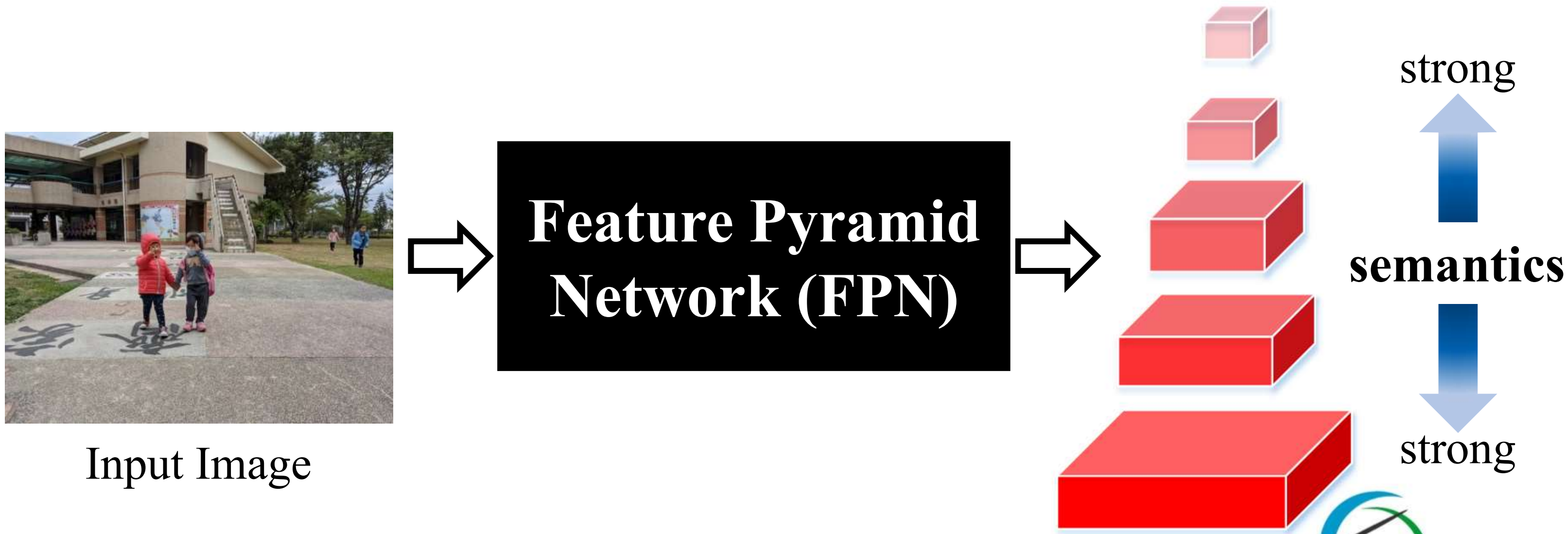
- Idea of FPN
  - **<u>propagate</u>** semantics from top layers (low-resolution) to bottom layers (high-resolution)

  ⇨ *all scales have rich semantics*



Network Architecture

- Overview
  - Input: an image with an arbitrary size
  - Output: a feature pyramid that all scales have rich semantics

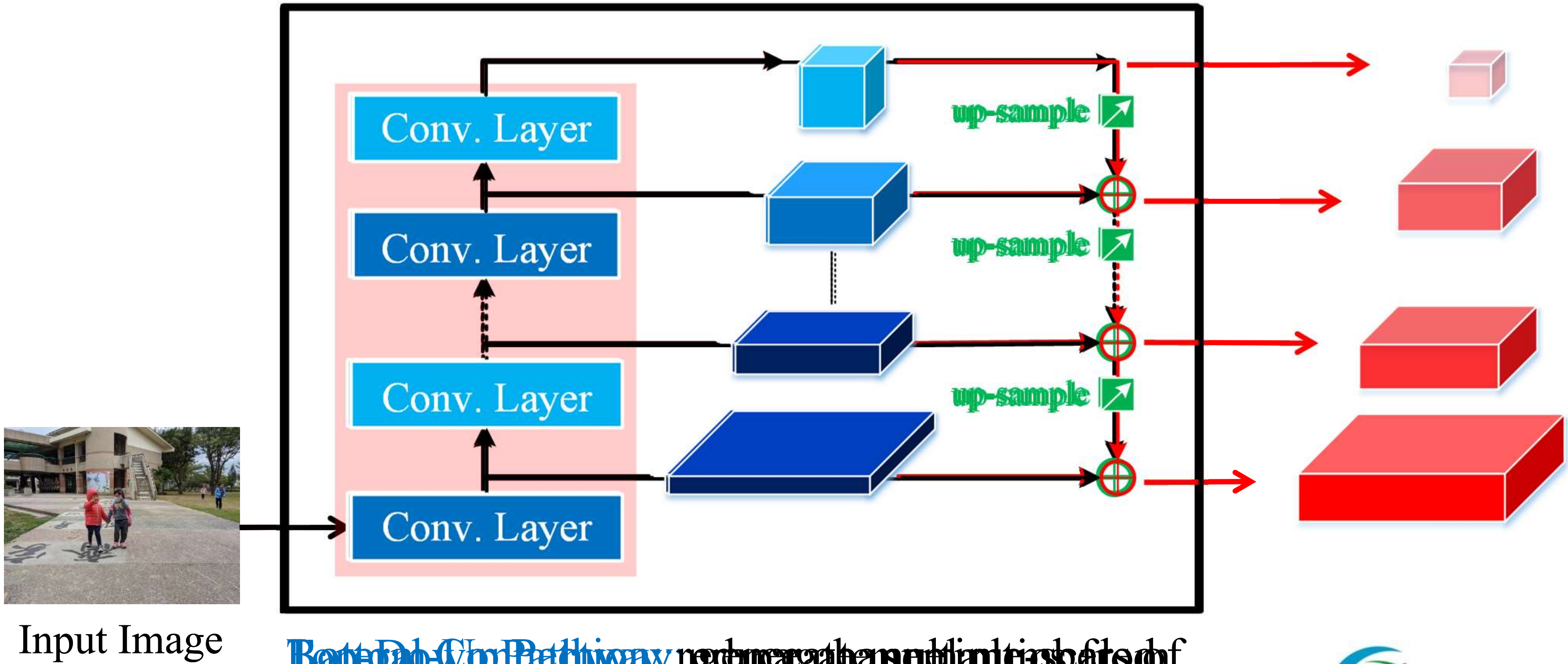


Network Architecture

- Overview
  - Bottom-Up Pathway: generate multiple-scaled feature maps in a pyramidal shape
  - Lateral Connection: reduce channel numbers of all feature maps to a fixed size for merging.
  - Top-Down Pathway: propagate semantics from topmost feature map to bottommost one.

Network Architecture

Feature Pyramid Network (FPN)

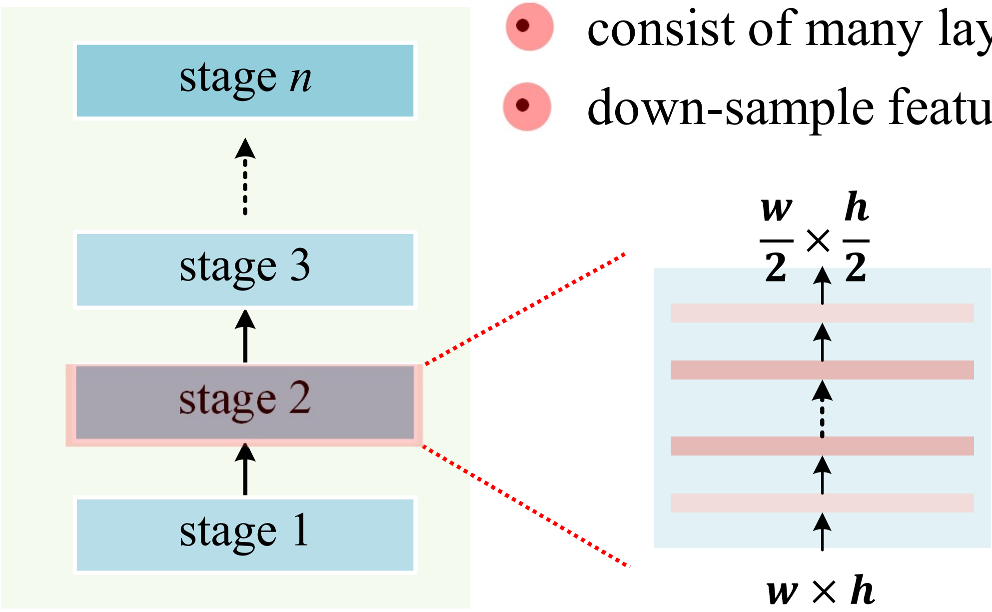


[illegible]

Network Architecture

- Bottom-Up Pathway
  - be a backbone Conv. network composed of **<u>several stages</u>**

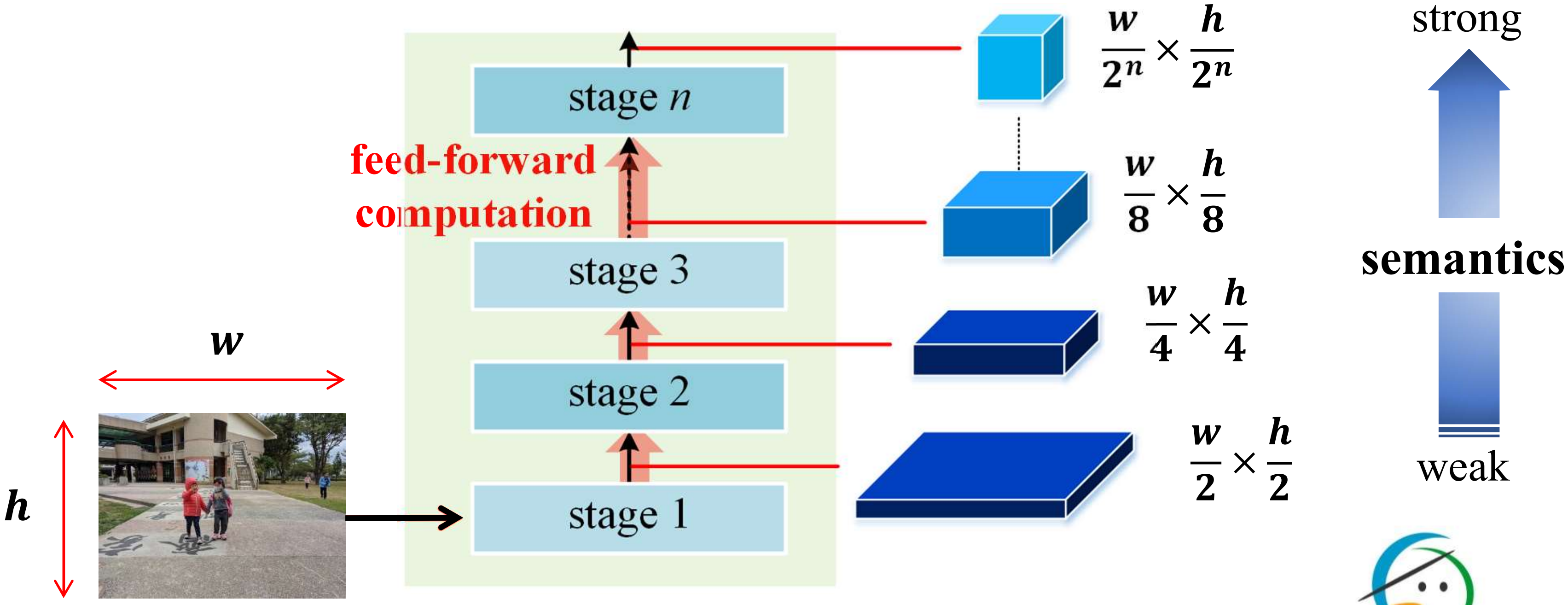
- consist of many layers
- down-sample feature map by a factor of 2



backbone Conv. network

Network Architecture

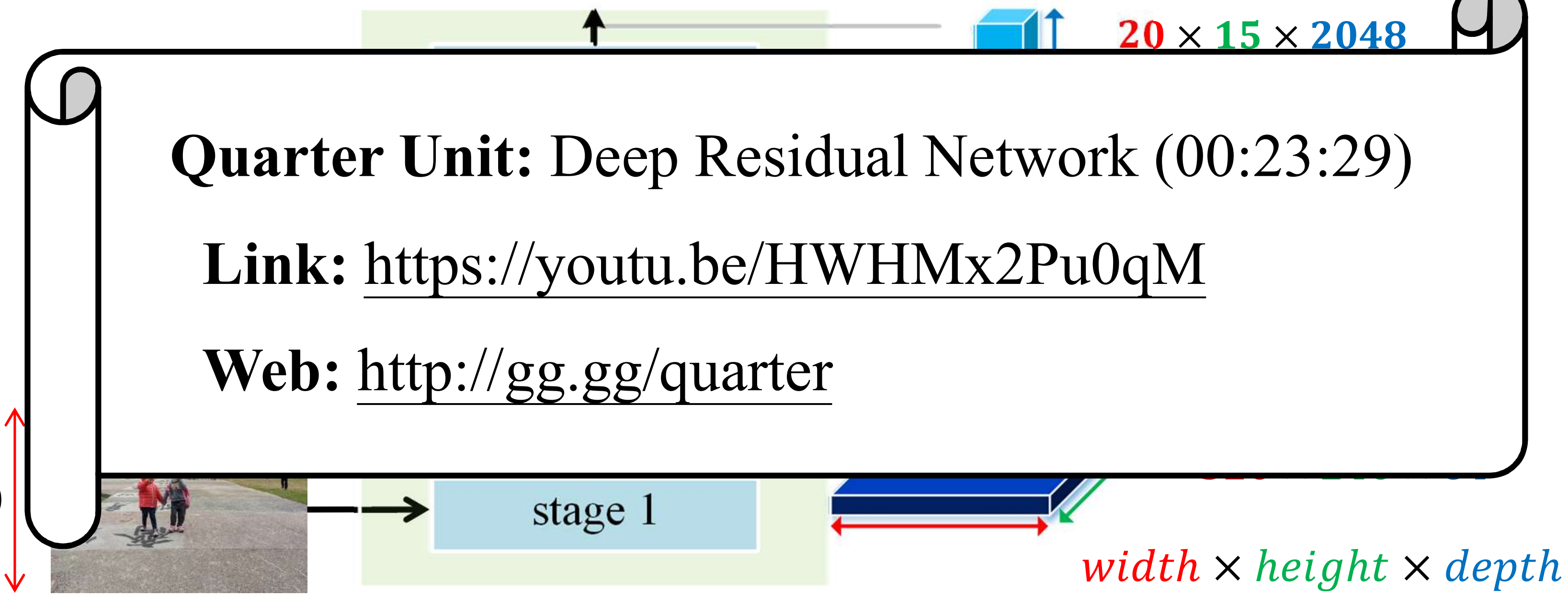
- Bottom-Up Pathway
  - perform feed-forward computation to produce a feature hierarchy with a scaling factor of 2



backbone Conv. network

Network Architecture

- Bottom-Up Pathway
  - example: **residual networks** consisting of 5 stages



Quarter Unit: Deep Residual Network (00:23:29)

Link: https://youtu.be/HWHMx2Pu0qM

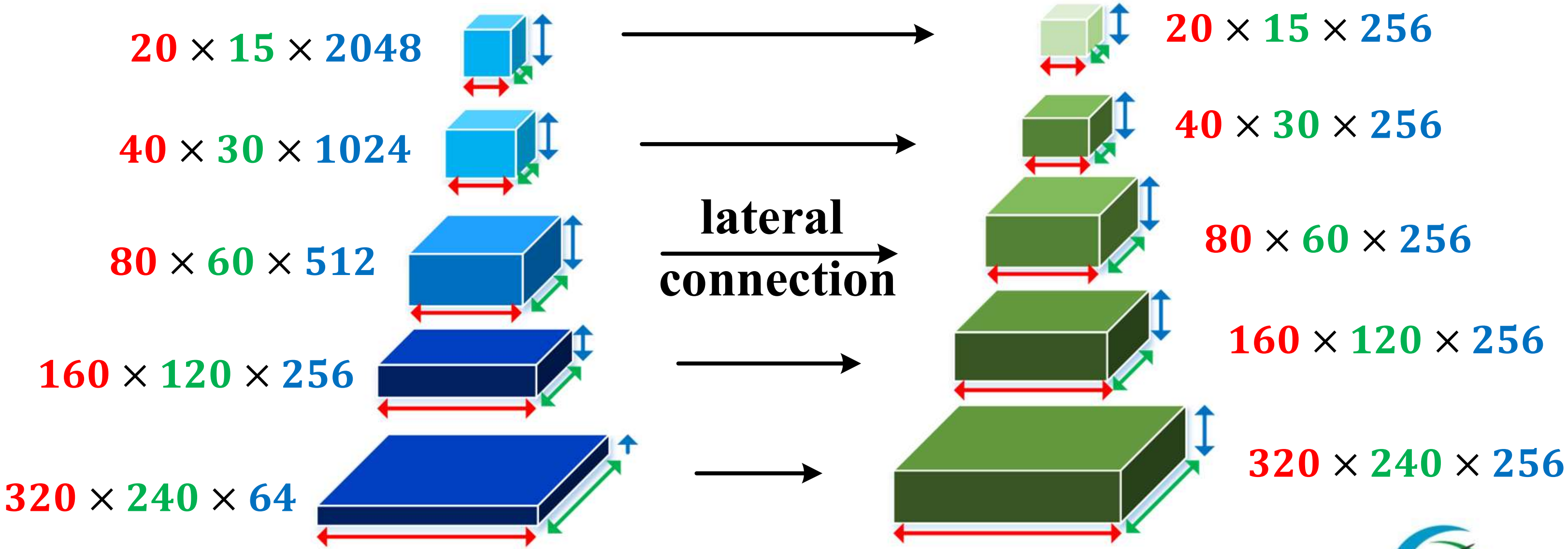
Web: http://gg.gg/quarter

residual networks

Kaiming He, *et. al.*, "Deep Residual Learning for Image Classification", CVPR 2016

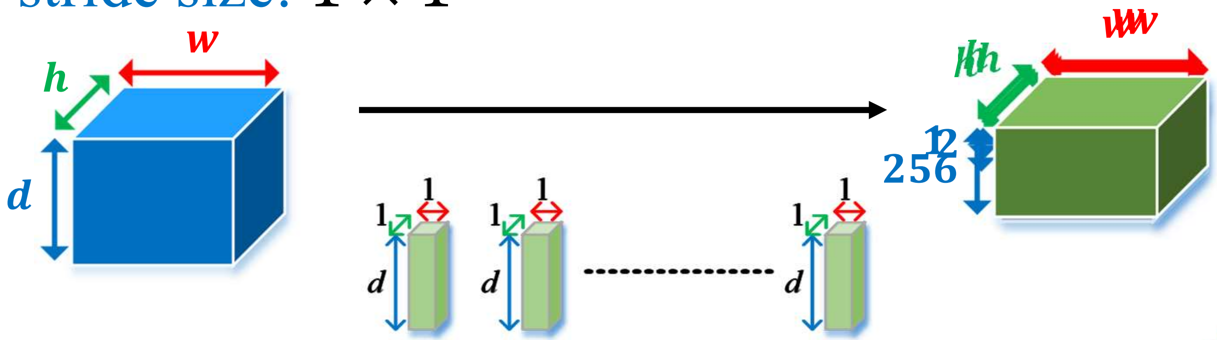
Network Architecture

- Lateral Connection
  - reduce channel numbers of all pyramidal feature maps to a fixed value (i.e. 256)



Network Architecture

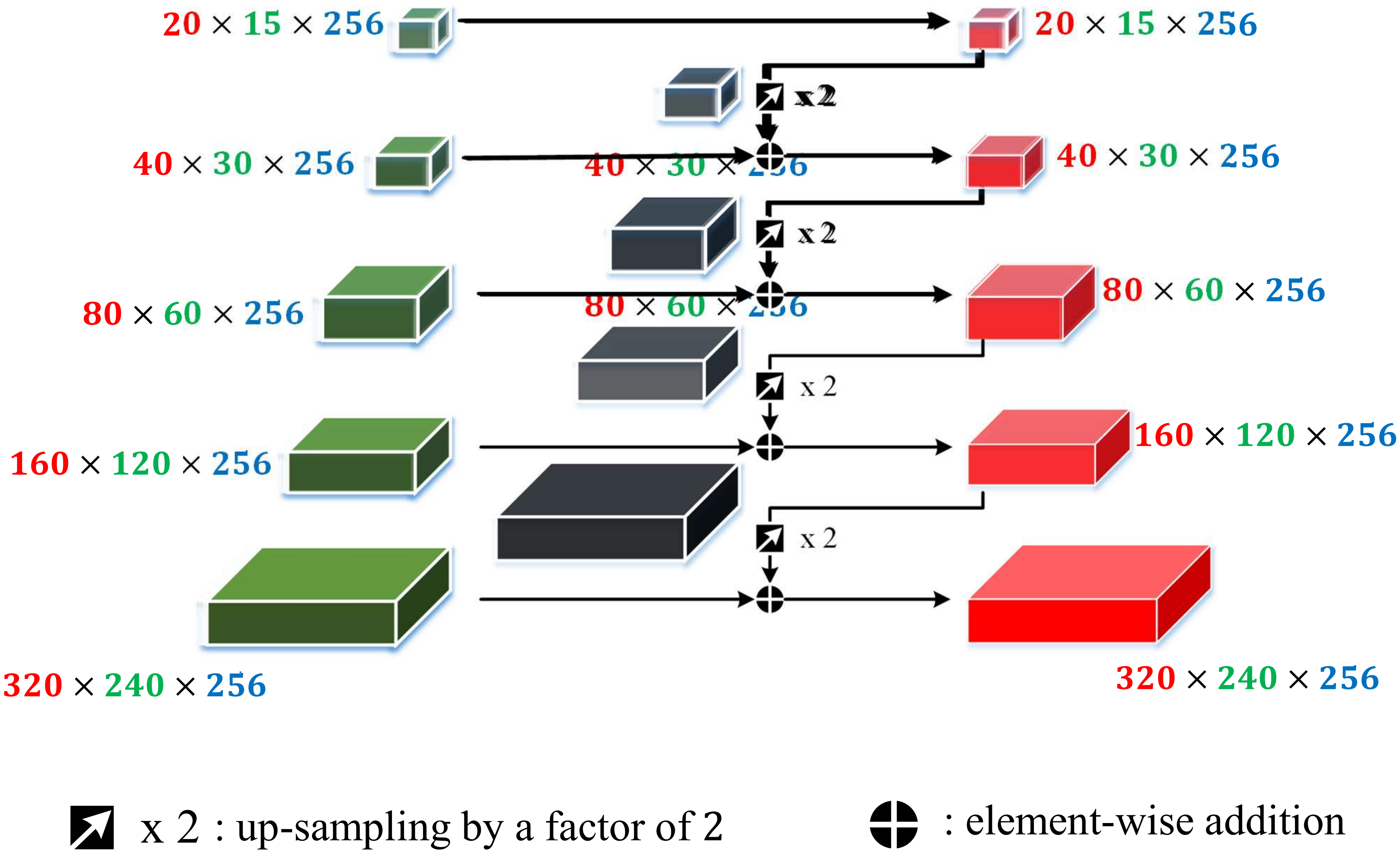
- Lateral Connection
  - perform channel reduction of all feature maps via convolution layers
    - kernel no.: 256
    - kernel size: $1 \times 1 \times d$
    - stride size: 1×1



256 kernels

Network Architecture

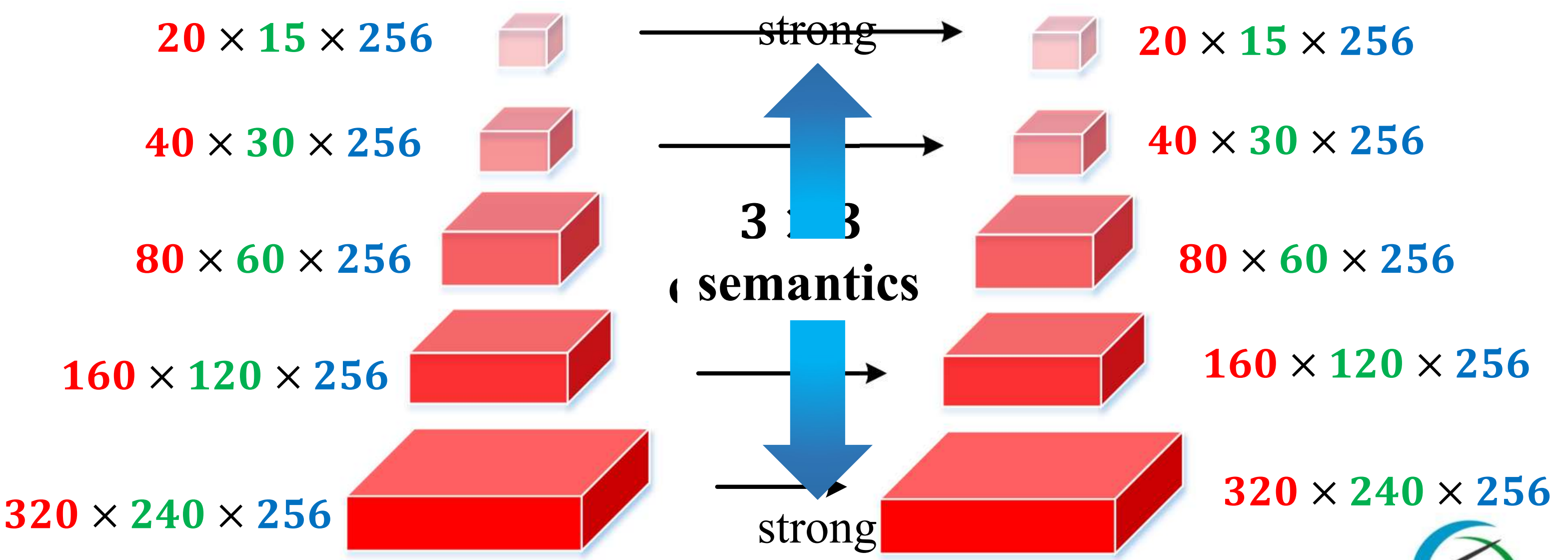
- Top-Down Pathway
  - propagate **semantics** from the topmost feature map to the bottommost one
    - up-sample the upper feature map by a factor of 2 using **nearest neighbor strategy**
    - merge the up-sampled feature map and lower one by **element-wise addition**.



20 × 15 × 256
40 × 30 × 256
80 × 60 × 256
160 × 120 × 256
320 × 240 × 256
x2
x2
x 2
x 2
20 × 15 × 256
40 × 30 × 256
80 × 60 × 256
160 × 120 × 256
320 × 240 × 256
x 2 : up-sampling by a factor of 2
: element-wise addition

Network Architecture

- Top-Down Pathway
  - reduce **aliasing effect** from up-sampling by a 3×3 convolution (padding size = 1)

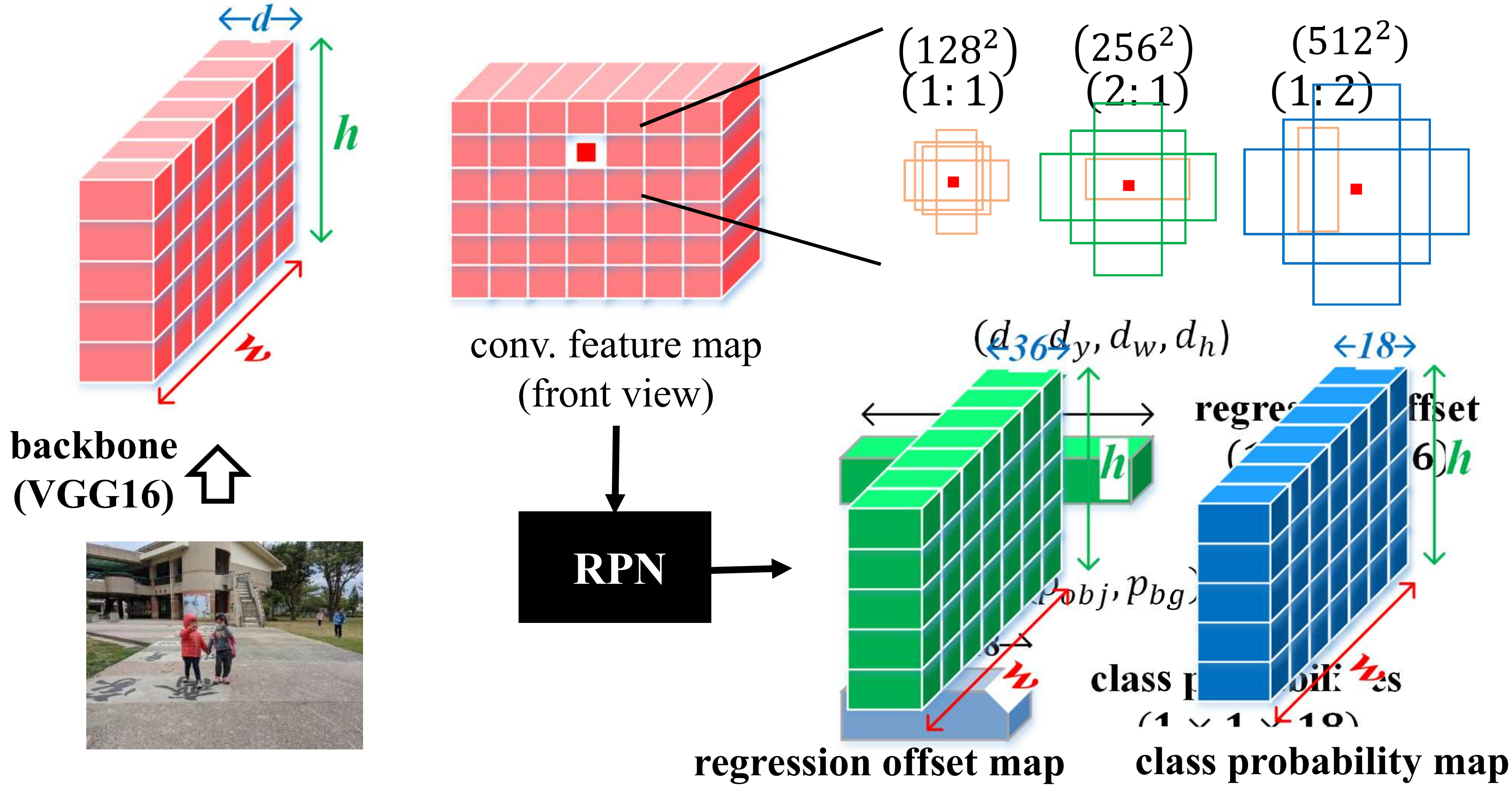


Region Proposal Network (RPN)

- About RPN
  - RPN is a network widely used in two-stage object detection.
  - RPN generates object proposals by using **dense anchor mechanism**.
    - attach 9 anchors (3 aspect ratios × 3 scales) centered at each point of the conv. feature map
    - predict one proposal with 6 parameters (4 regression + 2 class probabilities) w.r.t. each anchor

S. Ren, et. al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Network", *NIPS,* 2015

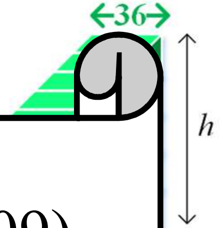
Region Proposal Network (RPN)



K. Simonyan et. al. "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ICLR*, 2015.

Region Proposal Network (RPN)

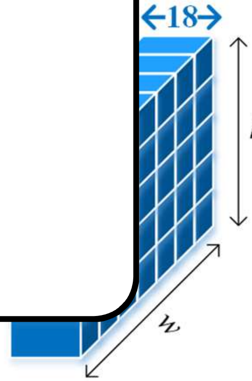
- About RPN



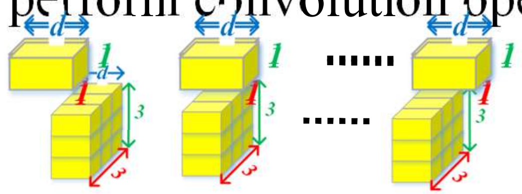
Quarter Unit: Region Proposal Network (00:19:09)

Link: https://youtu.be/0tBhRfEzUWs

Web: http://gg.gg/quarter



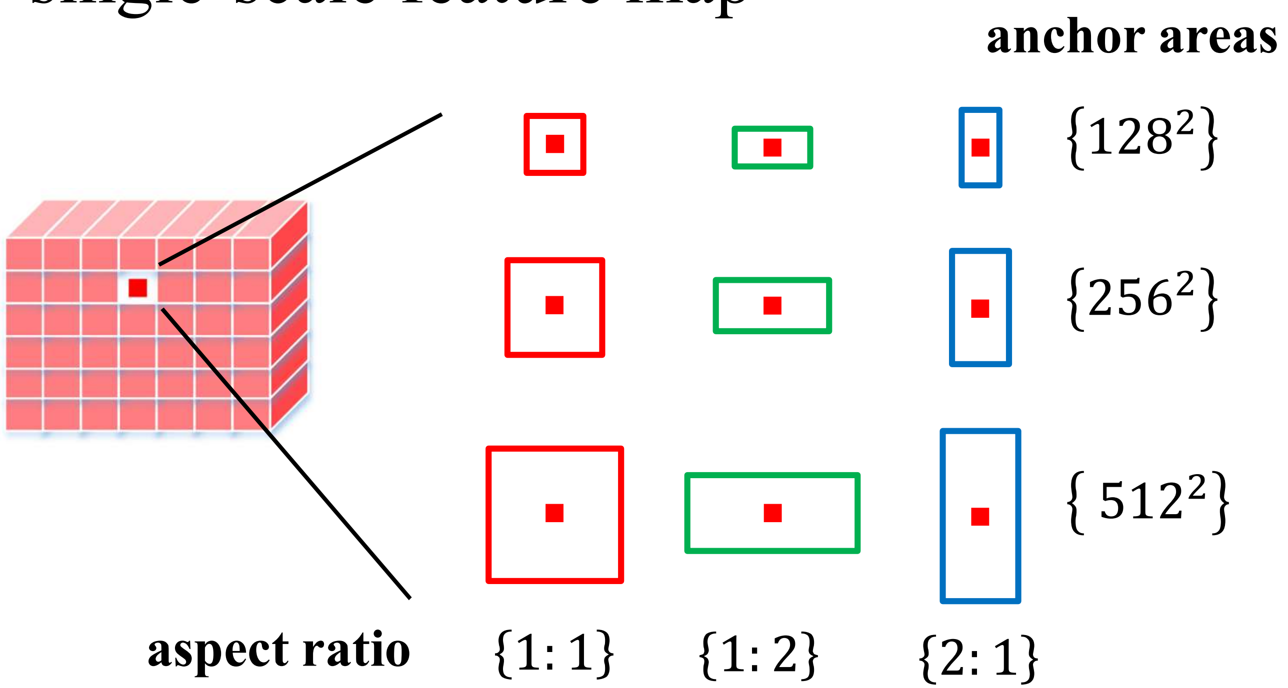
- predict the box parameters of all proposals
- predict the object/background probabilities of all proposals
- perform convolution operation



- kernel size: $1 \times 1 \times d$
- no. kernels: $36\ (= 9 \times 4)$
- kernel size: $3 \times 3 \times d$
- no. kernels: $18\ (= 9 \times 2)$
- no. kernels: d

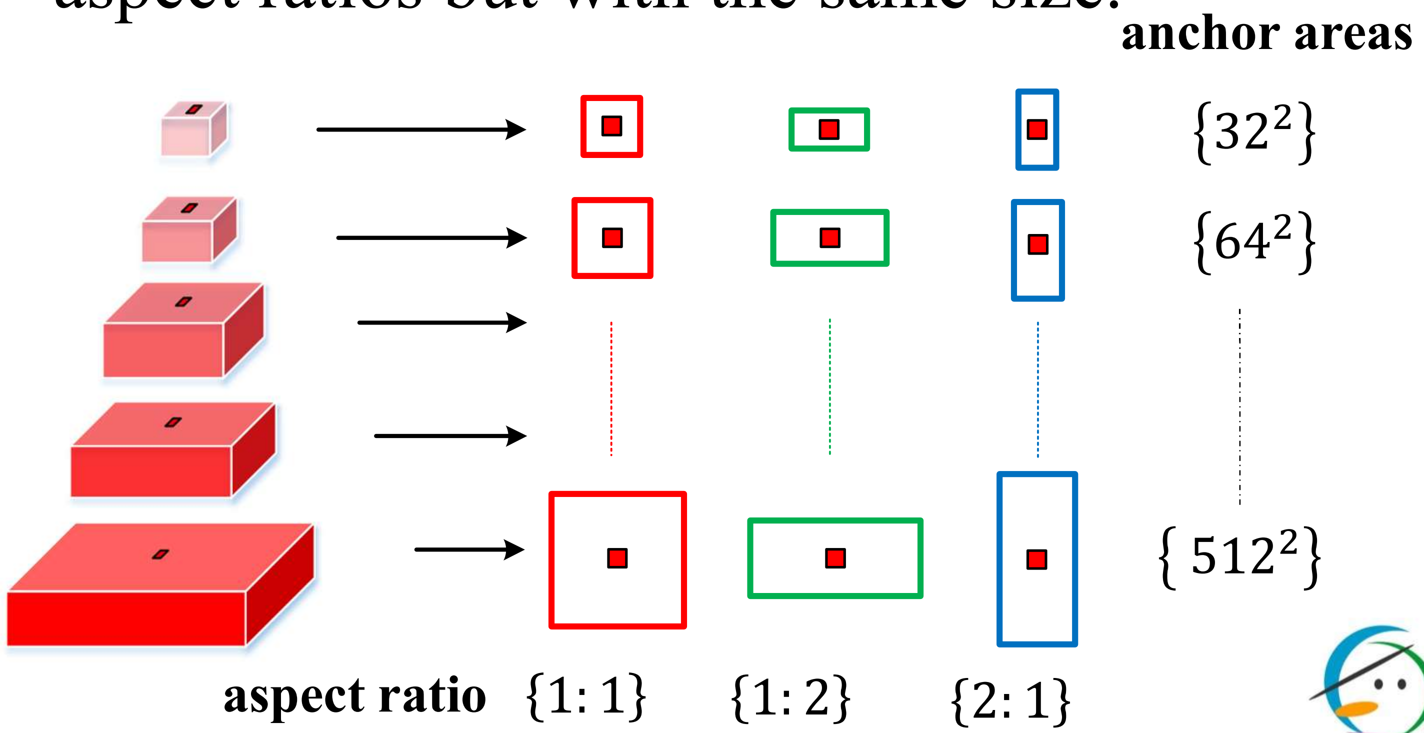
Region Proposal Network (RPN)

- Adopting FPN: anchor assignment
  - Basic RPN: attach 9 anchors to each point of a single-scale feature map



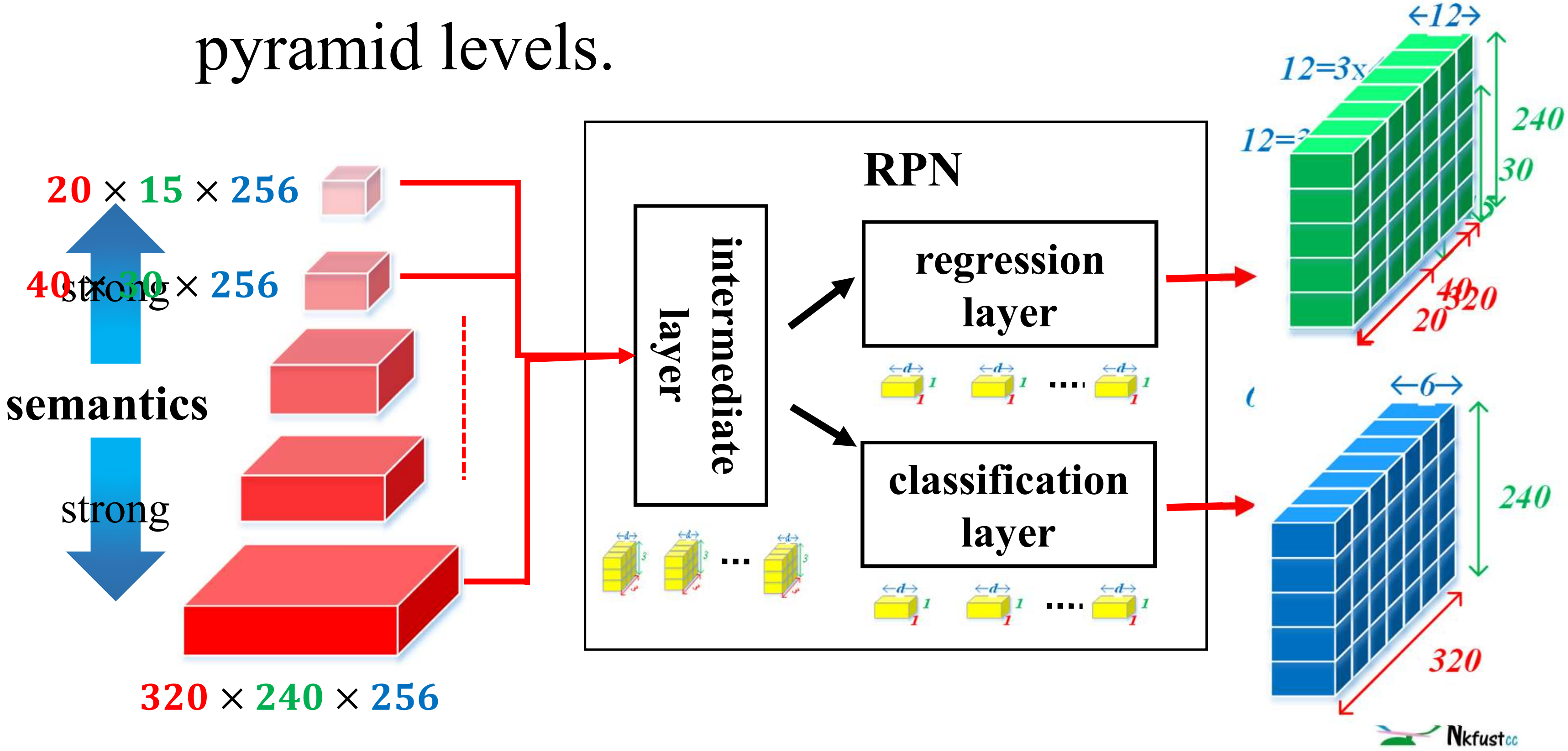
Region Proposal Network (RPN)

- Adopting FPN: anchor assignment
  - FPN+RPN: attach 3 anchors that are different in aspect ratios but with the same size.



Region Proposal Network (RPN)

- Adopting FPN
  - share the parameters of PRN across all feature pyramid levels.




thank you
© 1996, 2002 SANRIO CO., LTD.