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

Temporal Convolutional Network (TCN)

S. Bai, J. Z. Kolter, V. Koltun, 2018

Speaker: Shih-Shinh Huang

July 10, 2024

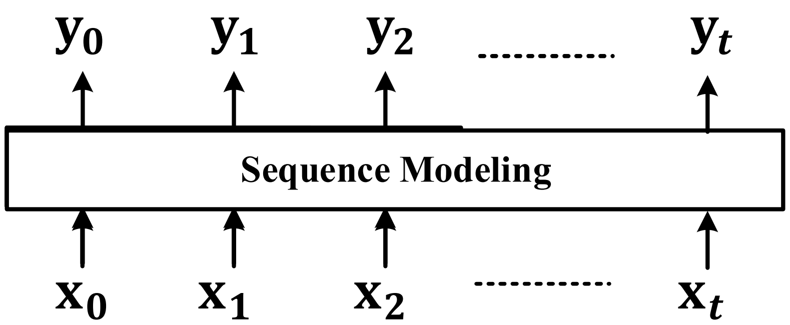
S. Bai, J. Z. Kolter, V. Koltun, , "An Empirical Evaluation of Generic Convolution and Recurrent Networks for Sequence Modeling", arXiv, 2018.

Outline

- Introduction
  - Motivation
  - Central Idea
  - Architecture Overview
- 1D Dilated Convolution
  - Definition
  - Padding Rule
- Residual Block
  - Dilated Conv. Layer
  - Residual Shortcut
  - Long-Memory Setting

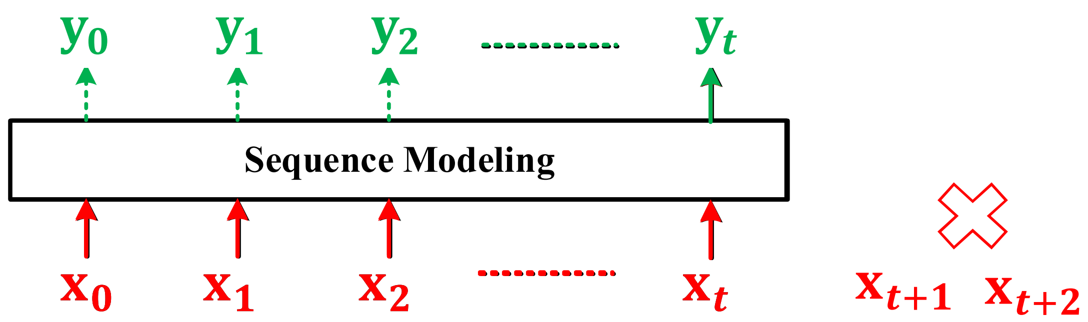
Introduction

- Motivation
  - **<u>sequence modeling</u>** is an essential task in many areas, such as weather/stock price prediction.
    - Given: an input sequence $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t$ to time t
    - Goal: predict the output $\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_t$ sequentially



Introduction

- Motivation
  - **<u>sequence modeling</u>** task has two properties
    - Causal Constraint: predicting $\mathbf{y}_t$ depends **<u>only on the past</u>** $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t$ and **<u>not on future</u>** $\mathbf{x}_{t+1}, \mathbf{x}_{t+2}, \ldots$.
    - Length Constraint: the length of the output sequence $\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_t$ is the same as the input $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t$



Introduction

- Motivation
  - **recurrent networks** are considered as the go-to solution for sequence modeling tasks.

Unit: Simple Recurrent Neural Network (RNN)

Link: https://youtu.be/Or9QSDqzOK0

Web: http://gg.gg/quarter

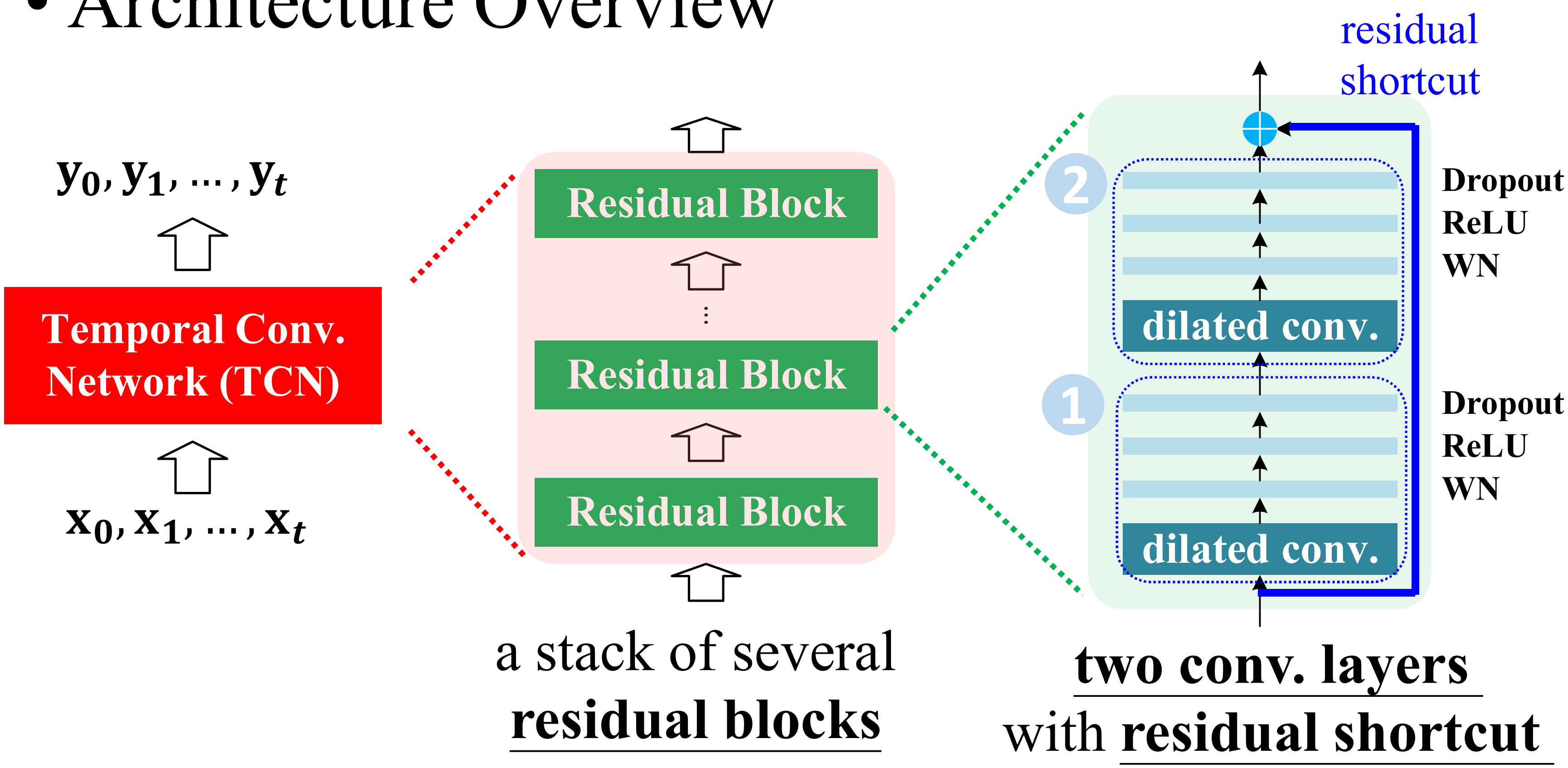
Introduction

- Central Idea
  - distill the design in the convolutional network to a simple network for sequence modeling.
    - 1D dilated convolution with zero padding: meet both the **causal** and **length constraints**
    - residual shortcut: allow for **very deep network** for extracting effective features.

- V. D. Oord, *et. al.*, "A Generative Model for Raw Audio," arXiv:1609,03499, 2016
- Kaiming He, *et. al.*, "Deep Residual Learning for Image Classification", CVPR 2016

Introduction

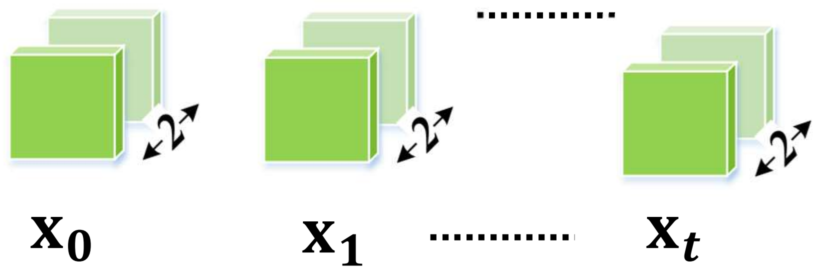
- Architecture Overview



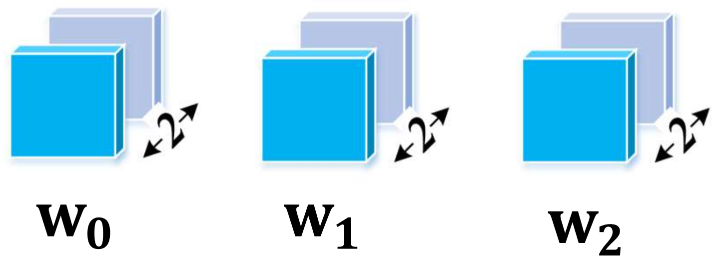
WN: weight normalization

1D Dilated Convolution

- Definition
  - $\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t\}$: input sequence with channel number = c (i.e. 2)



  - $\mathbf{W} = \{\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{k-1}\}$: kernel with size = k (i.e. 3) and channel number = c



1D Dilated Convolution

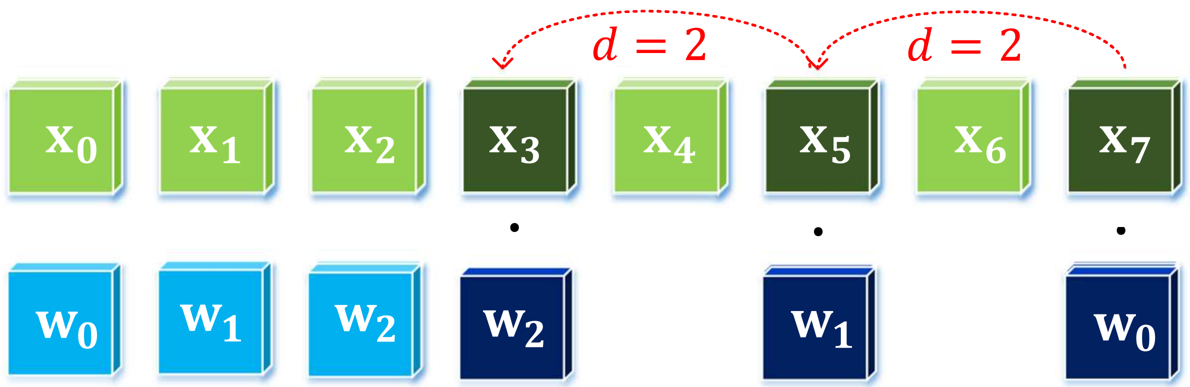
- Definition $(\mathbf{X} *_{d} \mathbf{W})$

inner product

$$k = 3 \quad d = 2 \;\Rightarrow\; (\mathbf{X} *_{d} \mathbf{W})(t) = \sum_{i=0}^{k-1} \mathbf{w}_i \cdot \mathbf{x}_{t-d\times i}$$

dilation factor

$$(\mathbf{X} * \mathbf{W})(t = 7) = \underline{\mathbf{w}_0 \cdot \mathbf{x}_{7-0}} + \underline{\mathbf{w}_1 \cdot \mathbf{x}_{7-2}} + \underline{\mathbf{w}_2 \cdot \mathbf{x}_{7-4}}$$



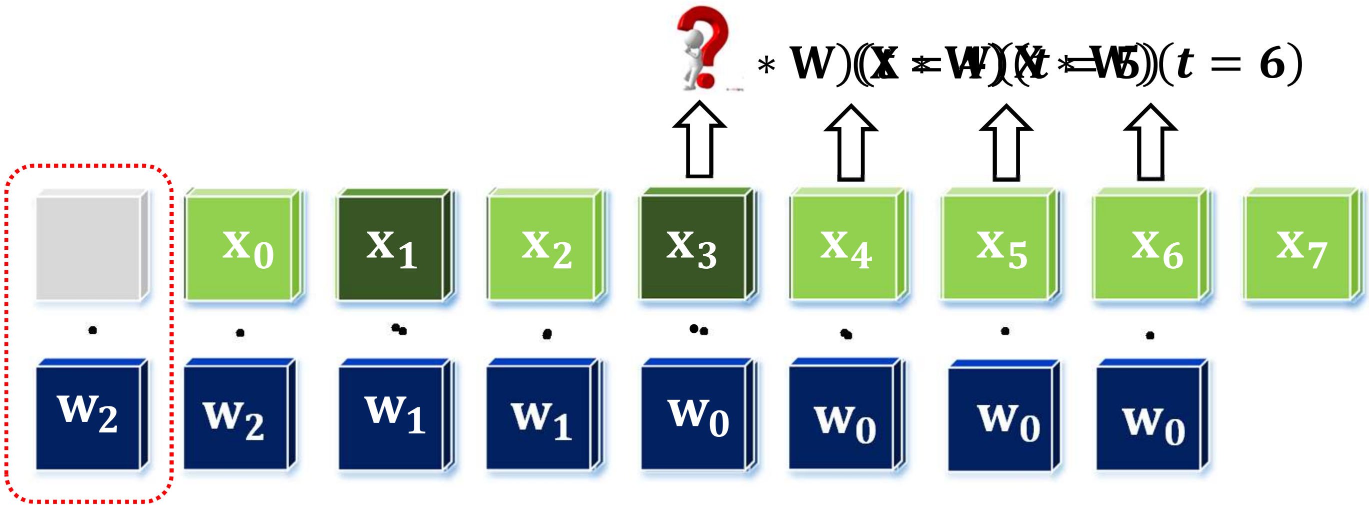
1D Dilated Convolution

- Definition $(\mathbf{X} *_d \mathbf{W})$

$$(\mathbf{X} * \mathbf{W})(t = 6) = \mathbf{w}_0 \cdot \mathbf{x}_{6-0} + \mathbf{w}_1 \cdot \mathbf{x}_{6-2} + \mathbf{w}_2 \cdot \mathbf{x}_{6-4}$$

$$(\mathbf{X} * \mathbf{W})(t = 5) = \mathbf{w}_0 \cdot \mathbf{x}_{5-0} + \mathbf{w}_1 \cdot \mathbf{x}_{5-2} + \mathbf{w}_2 \cdot \mathbf{x}_{5-4}$$

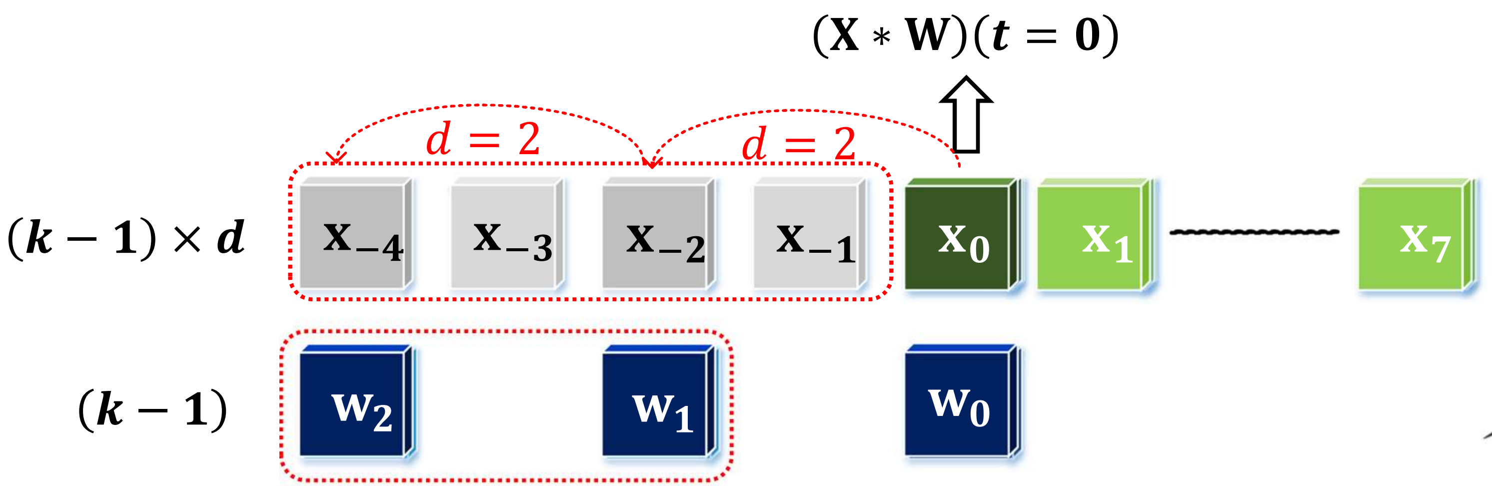
$$(\mathbf{X} * \mathbf{W})(t = 4) = \mathbf{w}_0 \cdot \mathbf{x}_{4-0} + \mathbf{w}_1 \cdot \mathbf{x}_{4-2} + \mathbf{w}_2 \cdot \mathbf{x}_{4-4}$$

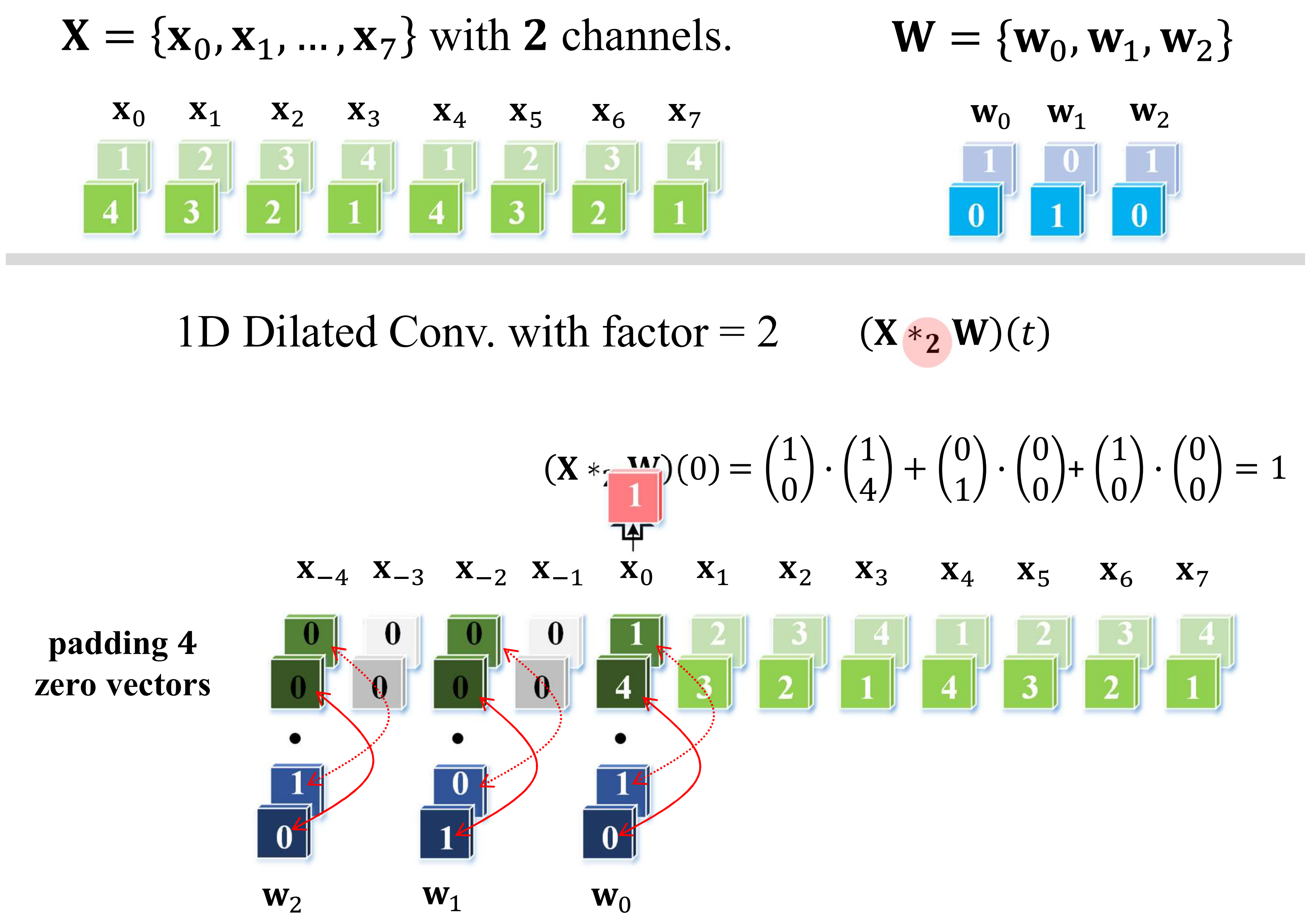


1D Dilated Convolution

- Padding Rule
  - add $\underline{(\boldsymbol{k}-\mathbf{1})\times \boldsymbol{d}}$ zero vectors to meet the **length constraint**

$$(\mathbf{X} * \mathbf{W})(\boldsymbol{t}=\mathbf{0}) = \mathbf{w_0}\cdot \mathbf{x_{0-0}} + \mathbf{w_1}\cdot \mathbf{x_{0-2}} + \mathbf{w_2}\cdot \mathbf{x_{0-4}}$$

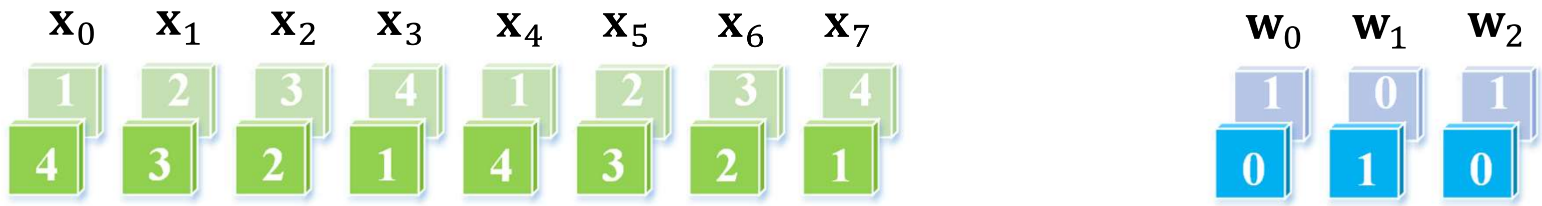




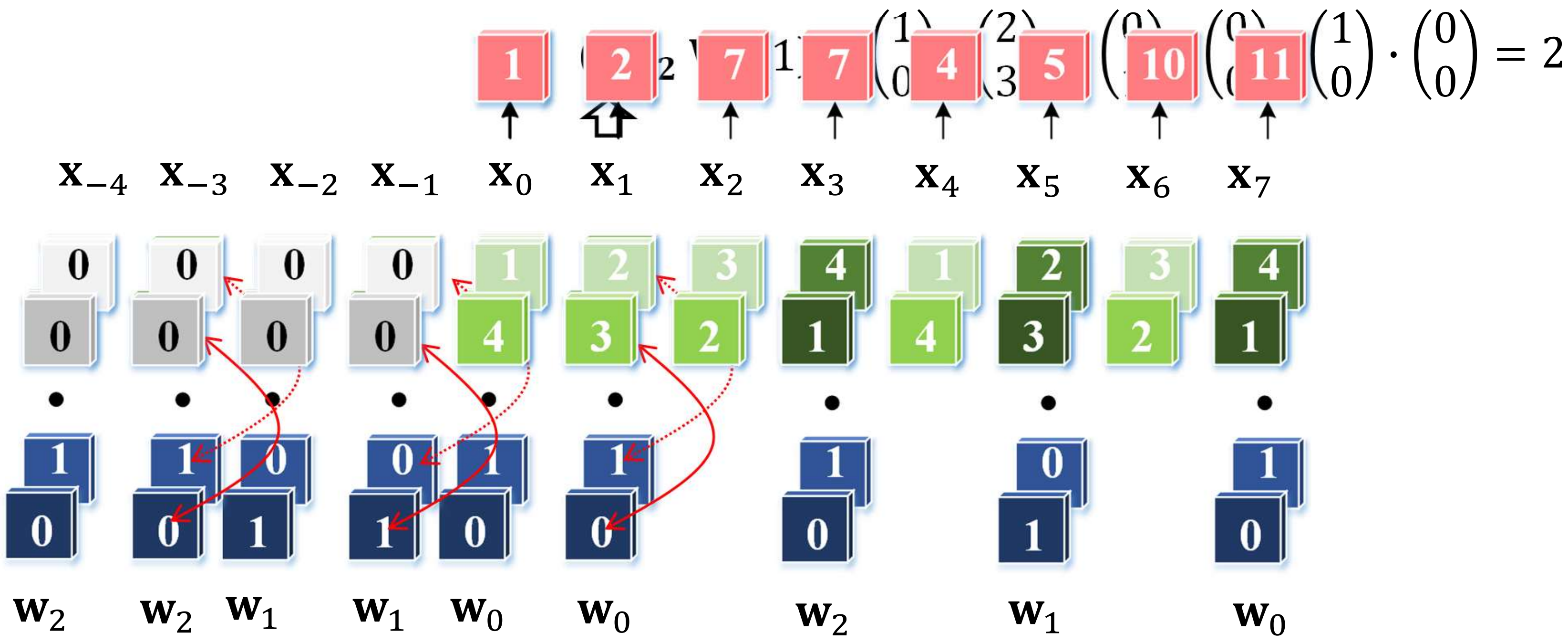
$\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_7\}$ with **2** channels.
$\mathbf{W} = \{\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2\}$
$\mathbf{x}_0$ $\mathbf{x}_1$ $\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_4$ $\mathbf{x}_5$ $\mathbf{x}_6$ $\mathbf{x}_7$
$\mathbf{w}_0$ $\mathbf{w}_1$ $\mathbf{w}_2$
1D Dilated Conv. with factor = 2
$(\mathbf{X} *_2 \mathbf{W})(t)$
$(\mathbf{X} *_2 \mathbf{W})(0) = \begin{pmatrix}1\\0\end{pmatrix} \cdot \begin{pmatrix}1\\4\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix} \cdot \begin{pmatrix}0\\0\end{pmatrix} + \begin{pmatrix}1\\0\end{pmatrix} \cdot \begin{pmatrix}0\\0\end{pmatrix} = 1$
$\mathbf{x}_{-4}$ $\mathbf{x}_{-3}$ $\mathbf{x}_{-2}$ $\mathbf{x}_{-1}$ $\mathbf{x}_0$ $\mathbf{x}_1$ $\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_4$ $\mathbf{x}_5$ $\mathbf{x}_6$ $\mathbf{x}_7$
padding 4
zero vectors
$\mathbf{w}_2$ $\mathbf{w}_1$ $\mathbf{w}_0$

$\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_7\}$ with **2** channels.

$\mathbf{W} = \{\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2\}$



1D Dilated Conv. with factor = 2 $(\mathbf{X} *_2 \mathbf{W})(t)$



Residual Block

approximate residual sequence $\mathbf{R} \approx \mathbf{Y} - \mathbf{X}$

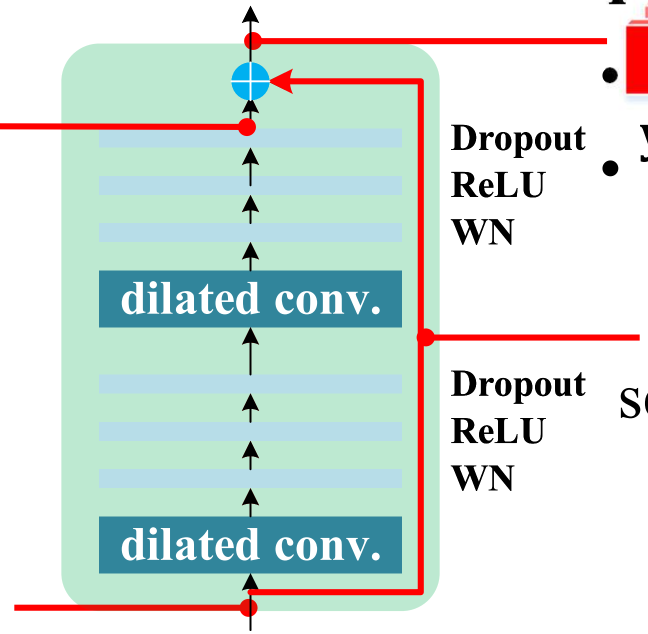
$\mathbf{R} = \{\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_7\}$

$\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{t-1}\}$

$c_{in} = 2$

- sequence len. = t
- no. channel = c_{in}

$\mathbf{x}_0$ $\mathbf{x}_1$ $\mathbf{x}_7$



$c_{out} = 3$

$\mathbf{Y} = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}$

- sequence len. = t
- no. channel = c_{out}

$\mathbf{y}_0$ $\mathbf{y}_1$ $\mathbf{y}_7$

resume target sequence $\mathbf{R} + \mathbf{X} \approx \mathbf{Y}$

Deep Residual Network

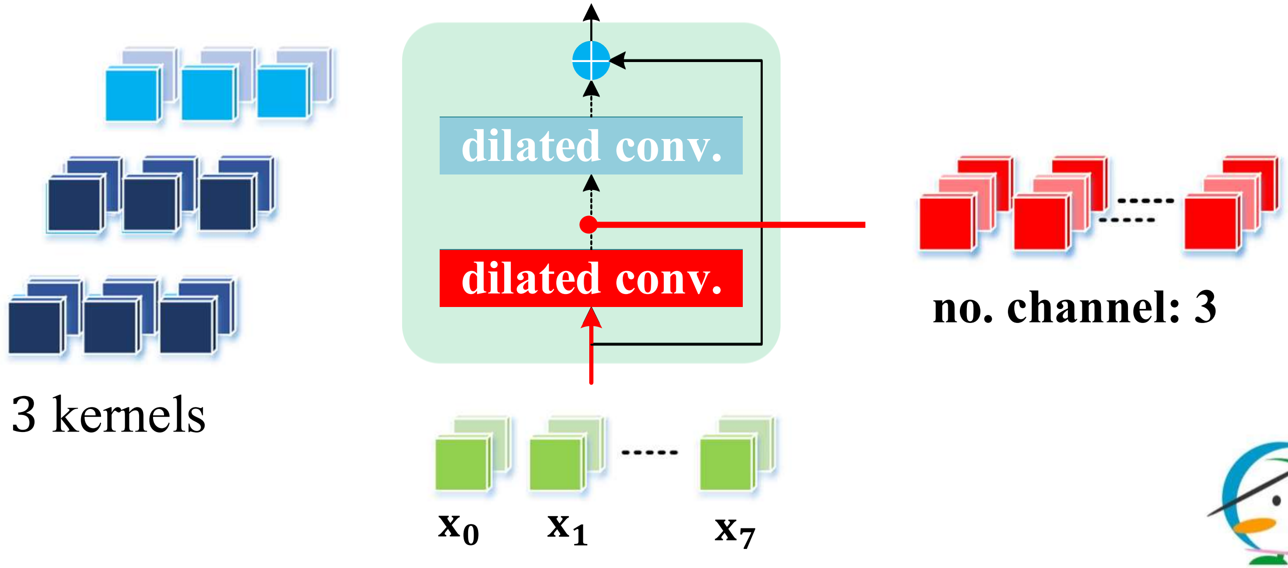
Link: https://youtu.be/HWHMx2Pu0qM

Web: http://gg.gg/quarter

Residual Block

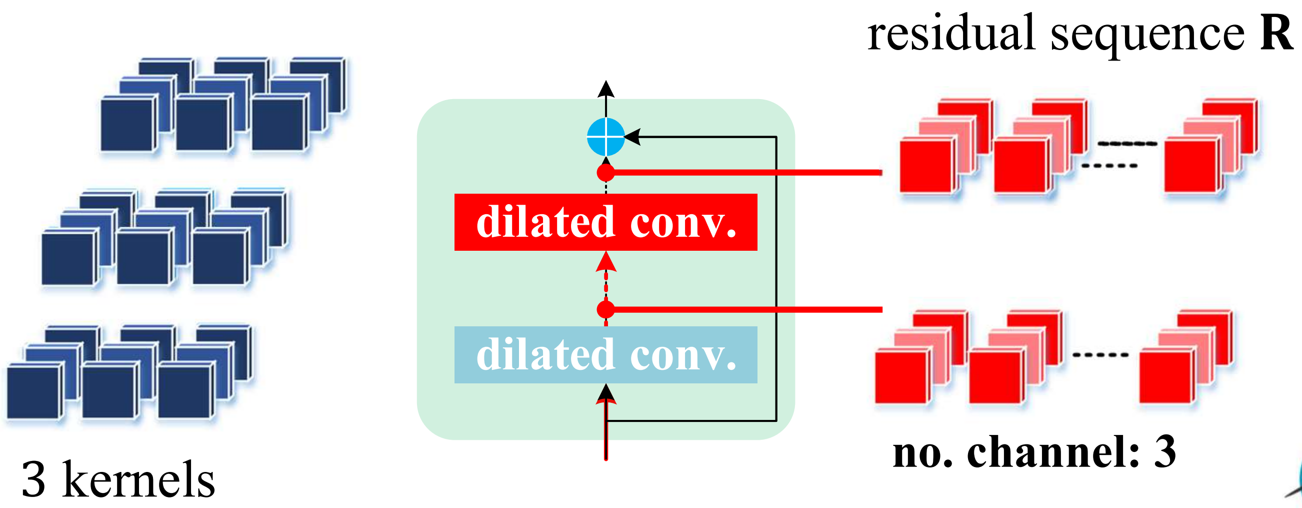
- Dilated Conv. Layer
  - convert no. channel $\boldsymbol{c_{in}}$ of $\mathbf{X}$ to $\boldsymbol{c_{out}}$
    - no. kernel: $c_{out}(= 3)$

ernel channel: $c_{in}(= 2)$



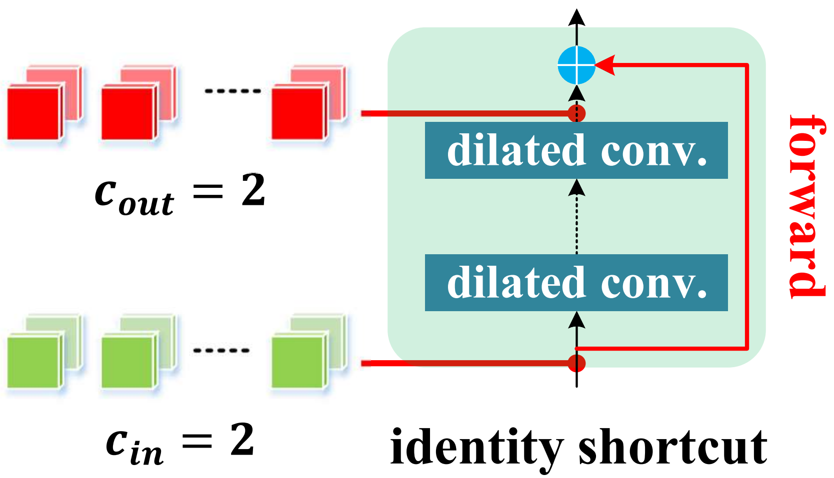
Residual Block

- Dilated Conv. Layer
  - output the residual sequence (no. channel: $\boldsymbol{c_{out}}$)
    - no. kernel: $c_{out}(=3)$
    - kernel channel: $c_{out}(=3)$



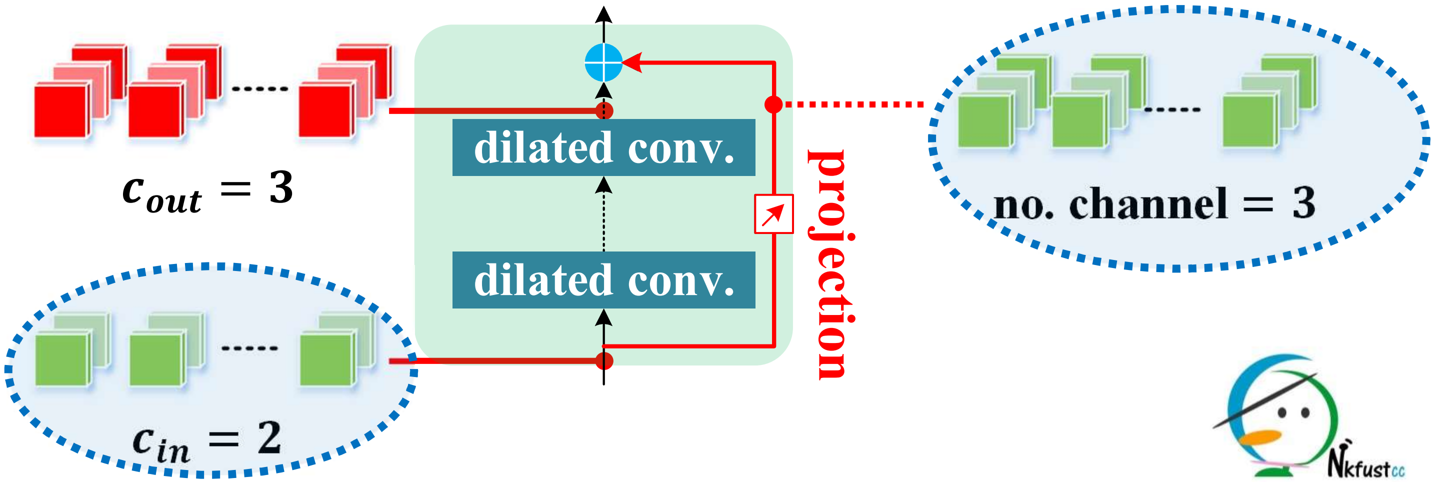
Residual Block

- Residual Shortcut
  - bypass the convolution operations
    - identity ($c_{in} = c_{out}$): directly forward input
    - projection ($c_{in} \neq c_{out}$): convert the input to have the same channel number as the output



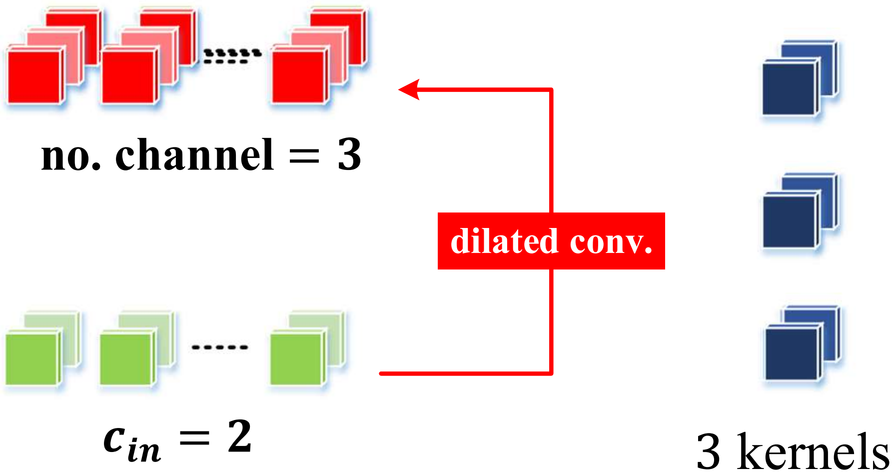
Residual Block

- Residual Shortcut
  - bypass the convolution operations
    - identity ($c_{in} = c_{out}$): directly forward input
    - projection ($c_{in} \neq c_{out}$): adjust the input to match the channel number of the output



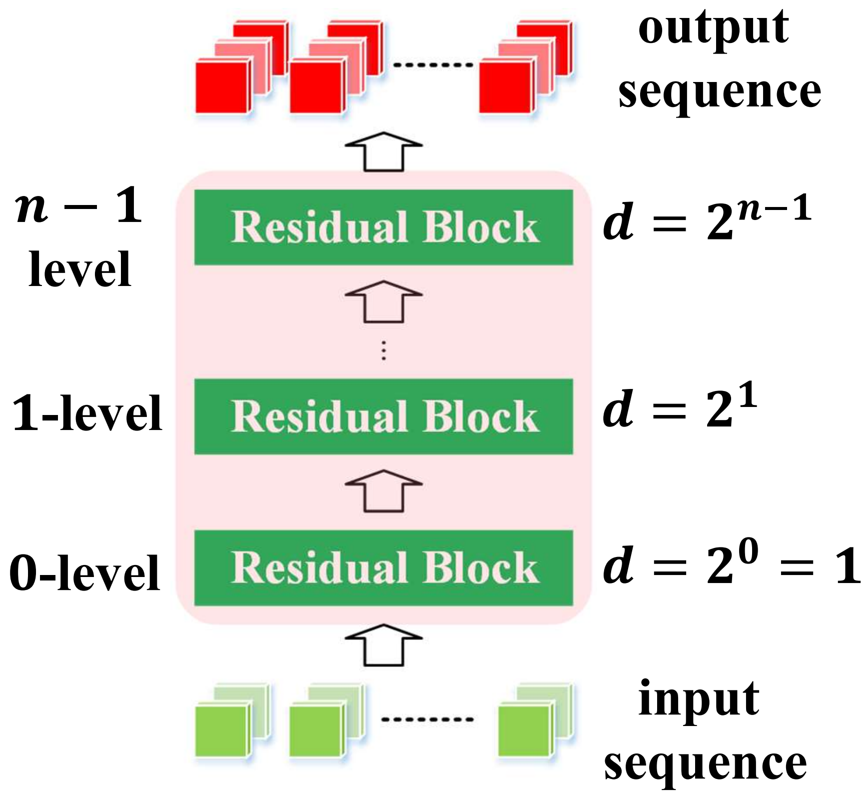
Residual Block

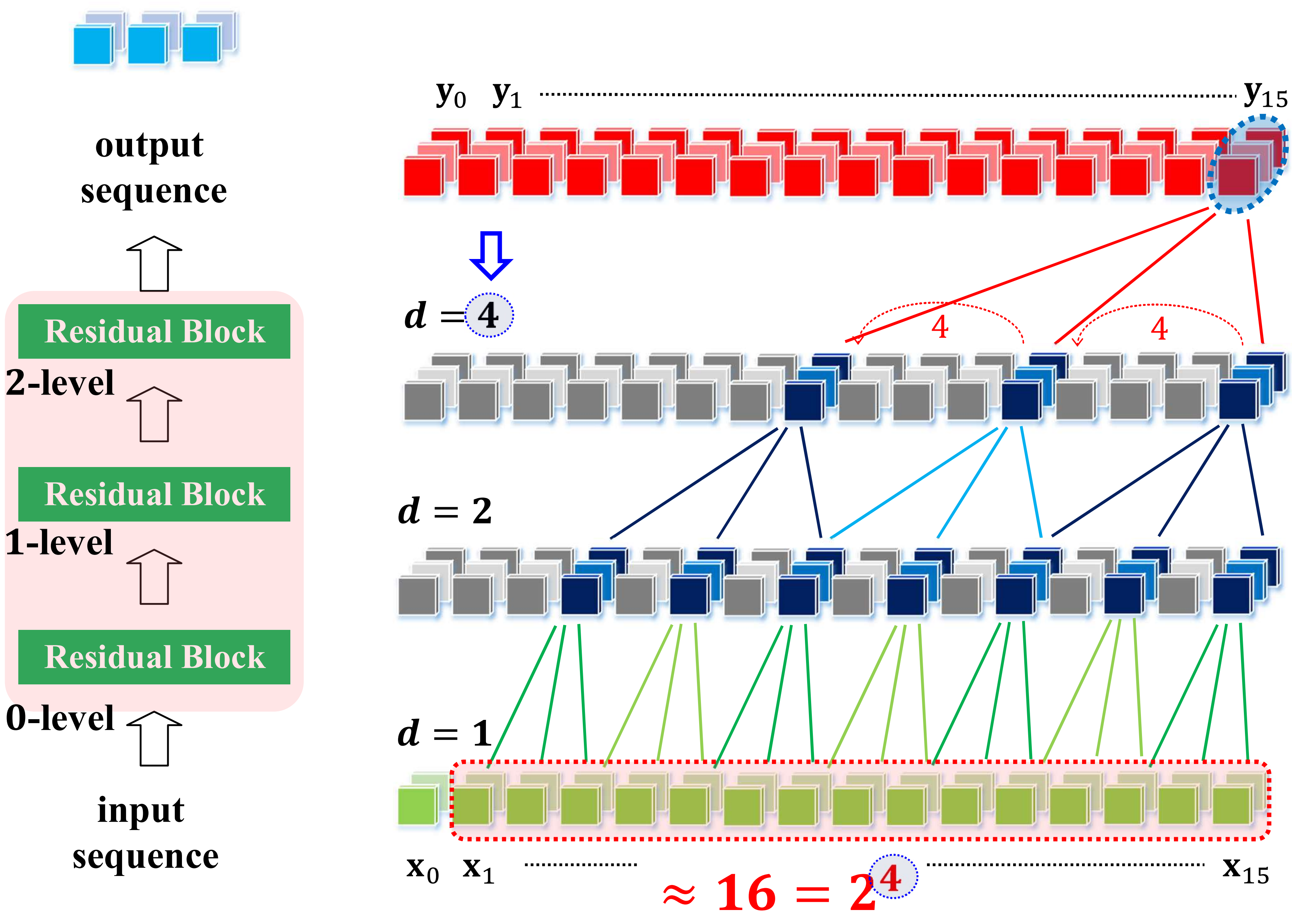
- Residual Shortcut
  - achieve channel mapping via 1D dilated conv.
    - no. kernel: $c_{out}(=3)$
    - kernel channel: $c_{in}(=2)$
    - kernel size: **1**



Residual Block

- Long-Memory Setting
  - make the network look very far into the past for prediction (long receptive field)
  - set the dilation factors d of residual blocks **<u>exponentially increasing</u>** relative to the depth of the network





output sequence
Residual Block
2-level
Residual Block
1-level
Residual Block
0-level
input sequence
y_0 y_1 y_{15}
$d = 4$
4
4
$d = 2$
$d = 1$
x_0 x_1 x_{15}
$\approx 16 = 2^4$


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