Wavenet for Text Generation

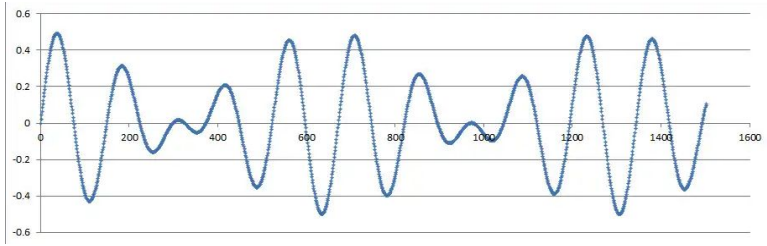
18. 10. 31

정 승 환

Sound Data vs NLP Data

- Sound Data vs NLP Data

Sound Data



$$f(x_t) = \text{sign}(x_t)\frac{\ln(1+\mu|x_t|)}{\ln(1+\mu)},$$

[0, 1, 3, 5, 8, 12, 17, …,13,11,3]

μ 값의 설정에 따라 index를 부여한 sequence 형태로 변환 가능

NLP Data

2017'(이하 ITU 2017)에서 '5G로 새로워지는 대한민국으로의 초대(Welcome to 5G Korea)'를 주제로 전시에 참가한다고 밝혔다.

오는 25일부터 나흘간 부산 벡스코에서 열리는 이번 ITU 2017에서 SK텔레콤은 400m2(약 121평) 규모의 전시관을 운영, 자율주행, 미디어, 인공지능, IoT(사물인터넷) 등 5개 영역에서 다양한 아이템을 선보인다.
SK텔레콤은 이번 전시에서 '16년부터 에릭슨·인텔과 공동 개발한 5G 이동형 인프라 차량을 처음 선보인다. 5G 이동형 인프라엔 5G 서비스를 제공하는데 필요한 모든 인프라와 서비스가 탑재됐다.

… <후략>

[1, 2, 7, 8, 9, 18, 203, …. 7, 9, 23, 24]

Sound Data vs NLP Data

- Sound Data vs NLP Data
  - Domain은 다르지만 기본적으로 index화 한 후의 데이터를 연속적으로 나열한 형태는 같다.
- 무엇이 더 어려운가..?
  - Sound의 경우에는 wave 형태이기 때문에 단조 증가 / 감소하는 형태를 가지지만 NLP에서는 그렇지 않다.
  - 하지만 Sound의 경우에는 상당히 긴 Sequence를 가지기 때문에 이 점은 Sound가 더 복잡하다.

Why Wavenet?

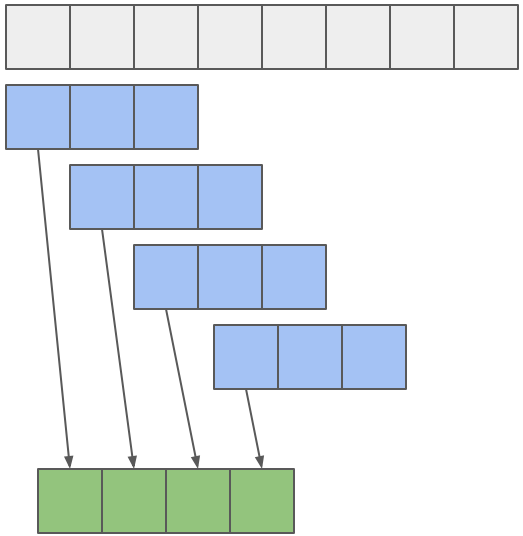
- RNN이 가지는 한계점
  - 상당히 긴 sequence를 다루기 어렵다.. (Sound Data의 경우에는 상당히 긴 Sequence를 가짐)
    - long-range dependency를 다루는데 한계를 가짐
  - multi-processing이 쉽지 않음
    - LSTM 의 경우 이전 stage의 hidden 값을 input으로 받기 때문에 parallel 하게 수행하기 쉽지 않음
- RNN에 대한 대안
  - CNN의 구조를 이용하여서 Sequence Data를 다루어 보자
  - WaveNet에서 CNN(Conv1D)를 이용하여 Sequence Data를 Input으로 하는 Network Architecture를 제시함

WaveNet

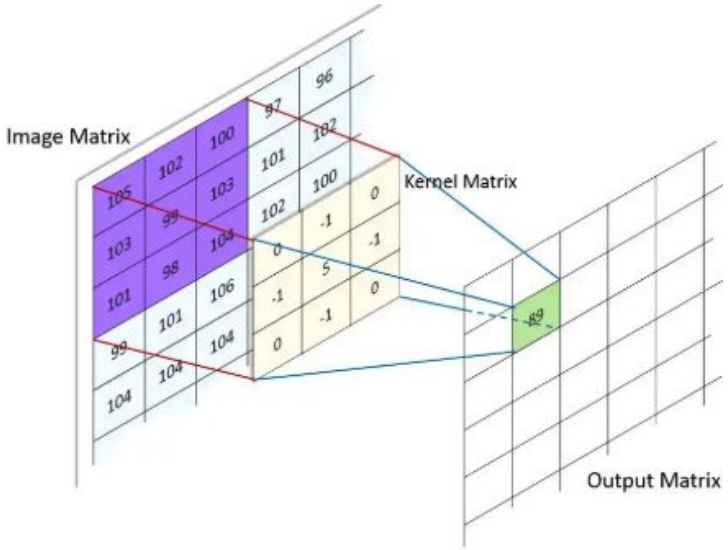
- 주요 용어들
  - Stack of casual convolution layers
  - Stack of casual convolutional layers
  - Stack of dilated casual convolution
  - Residual and skip connection
  - softmax distribution (음성 데이터 특성)

Convolution

- Conv 1D
  - 1 차원 데이터에 적용되는 Convolution



Conv 2D



https://www.quora.com/What-does-it-mean-by-1D-convolutional-neural-network
http://machinelearninguru.com/computer_vision/basics/convolution/image_convolution_1.html

Dilated convolution

- Dilated Convolutions in Conv2D

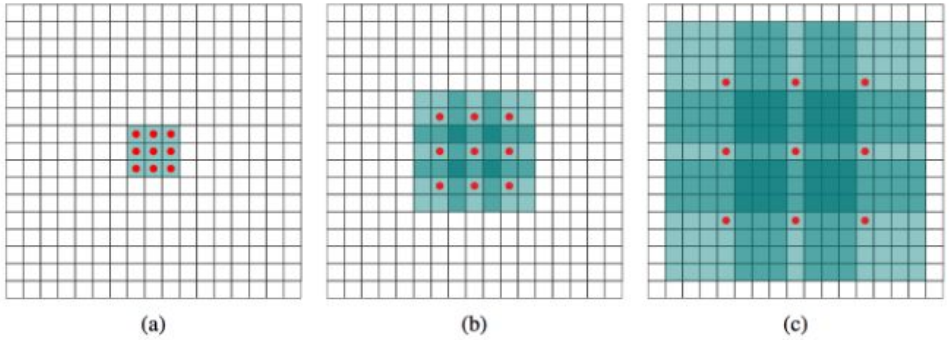


Figure 1: Systematic dilation supports exponential expansion of the receptive field without loss of resolution or coverage. (a) F_1 is produced from F_0 by a 1-dilated convolution; each element in F_1 has a receptive field of 3×3. (b) F_2 is produced from F_1 by a 2-dilated convolution; each element in F_2 has a receptive field of 7×7. (c) F_3 is produced from F_2 by a 4-dilated convolution; each element in F_3 has a receptive field of 15×15. The number of parameters associated with each layer is identical. The receptive field grows exponentially while the number of parameters grows linearly.

1	0	1	0	1
0	0	0	0	0
1	0	1	0	1
0	0	0	0	0
1	0	1	0	1

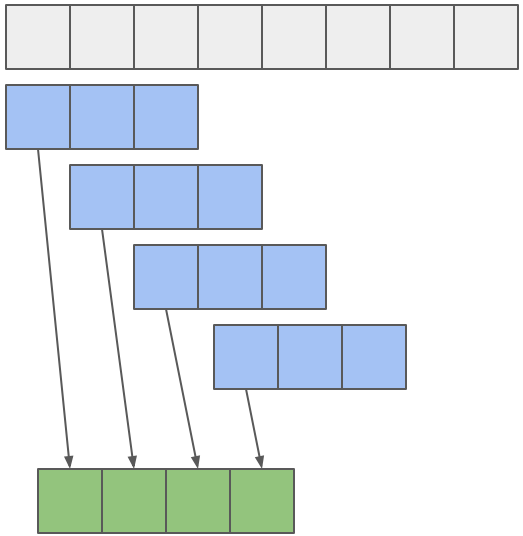
kernel size : 3 X 3
dilation : 2

http://sergeiturukin.com/2017/03/02/wavenet.html

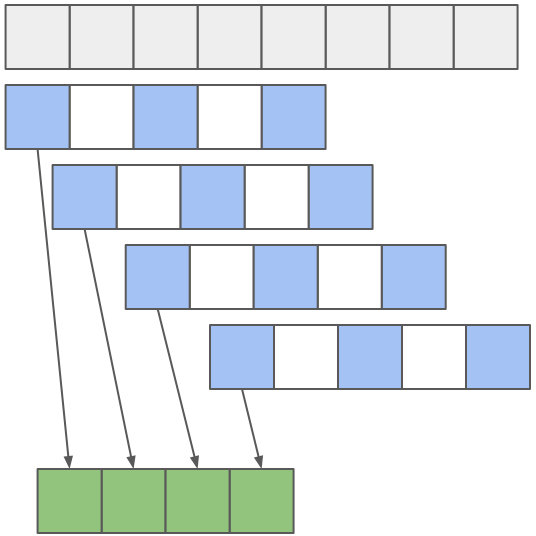
Stack of dilated casual convolution

- Dilated Conv 1D (kernel=3)

dilation =1



dilation = 2



http://sergeiturukin.com/2017/03/02/wavenet.html

Stack of casual convolution

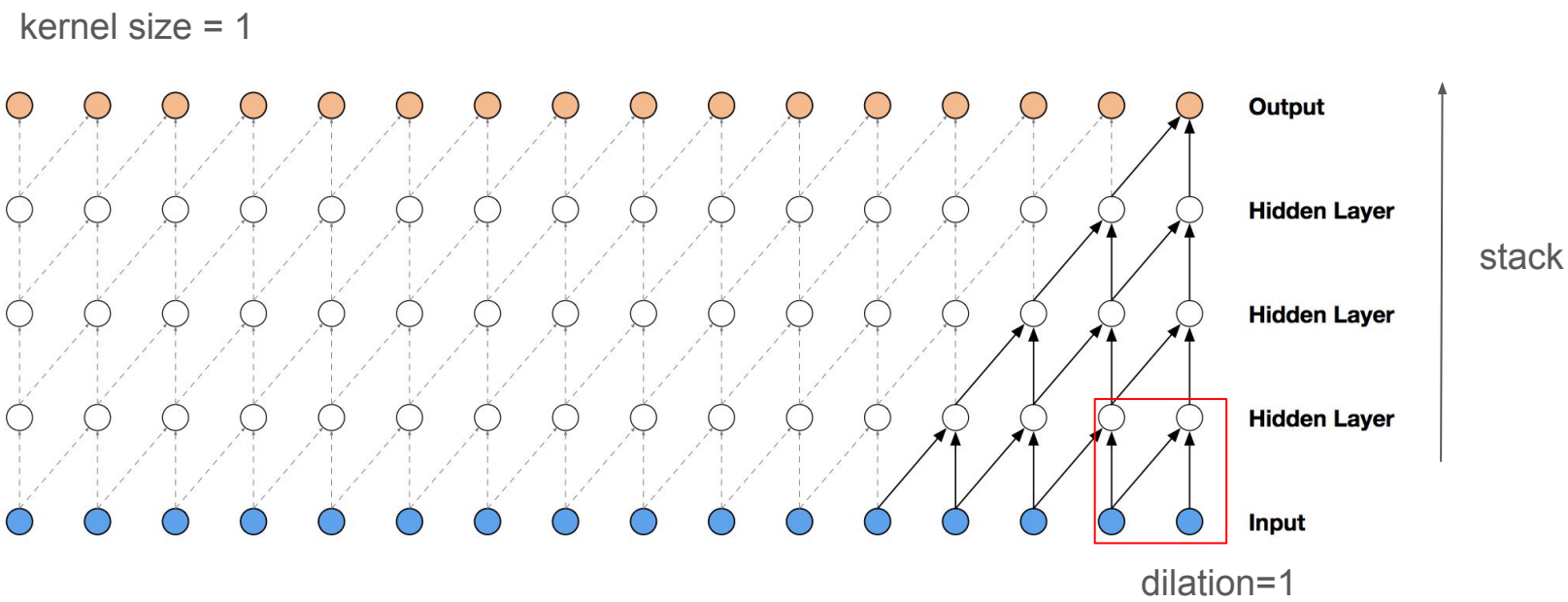


Figure 2: Visualization of a stack of causal convolutional layers.

https://arxiv.org/pdf/1609.03499.pdf

Stack of dilated casual convolution

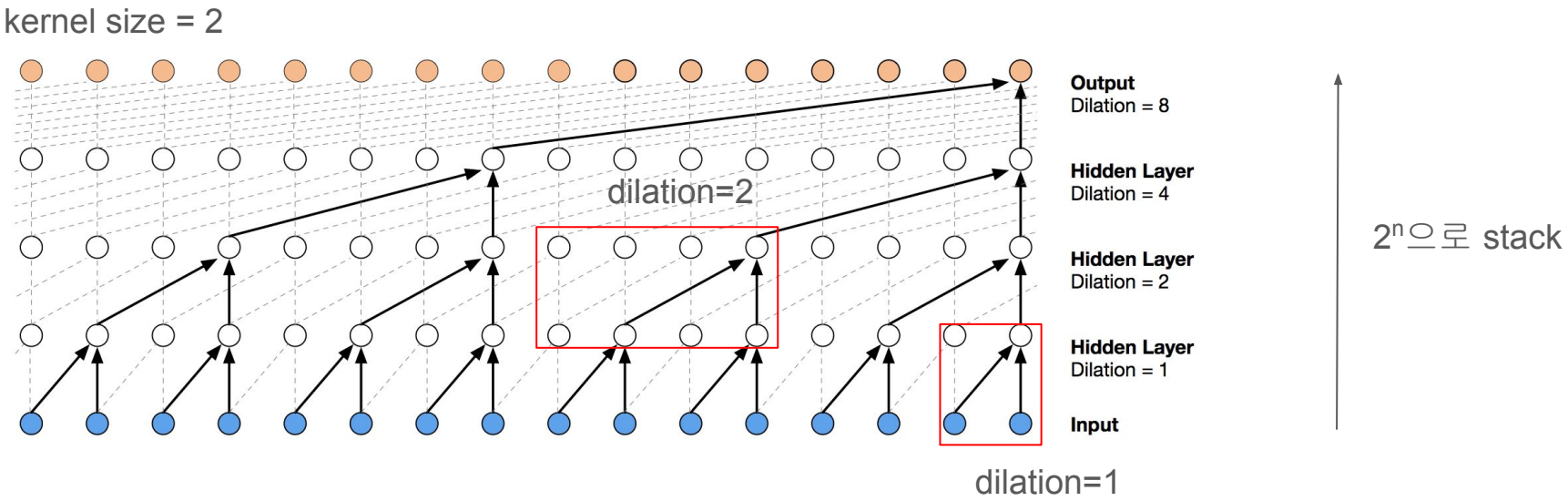
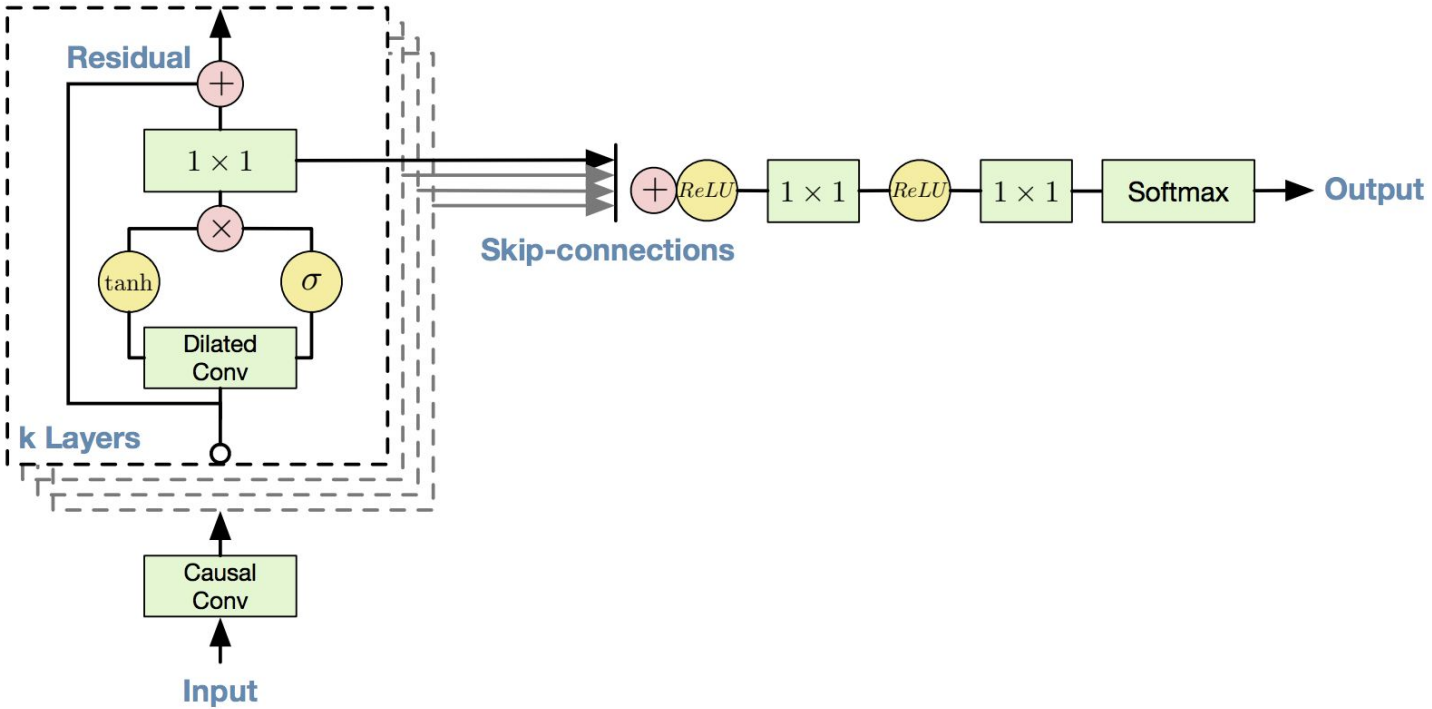


Figure 3: Visualization of a stack of *dilated* causal convolutional layers.

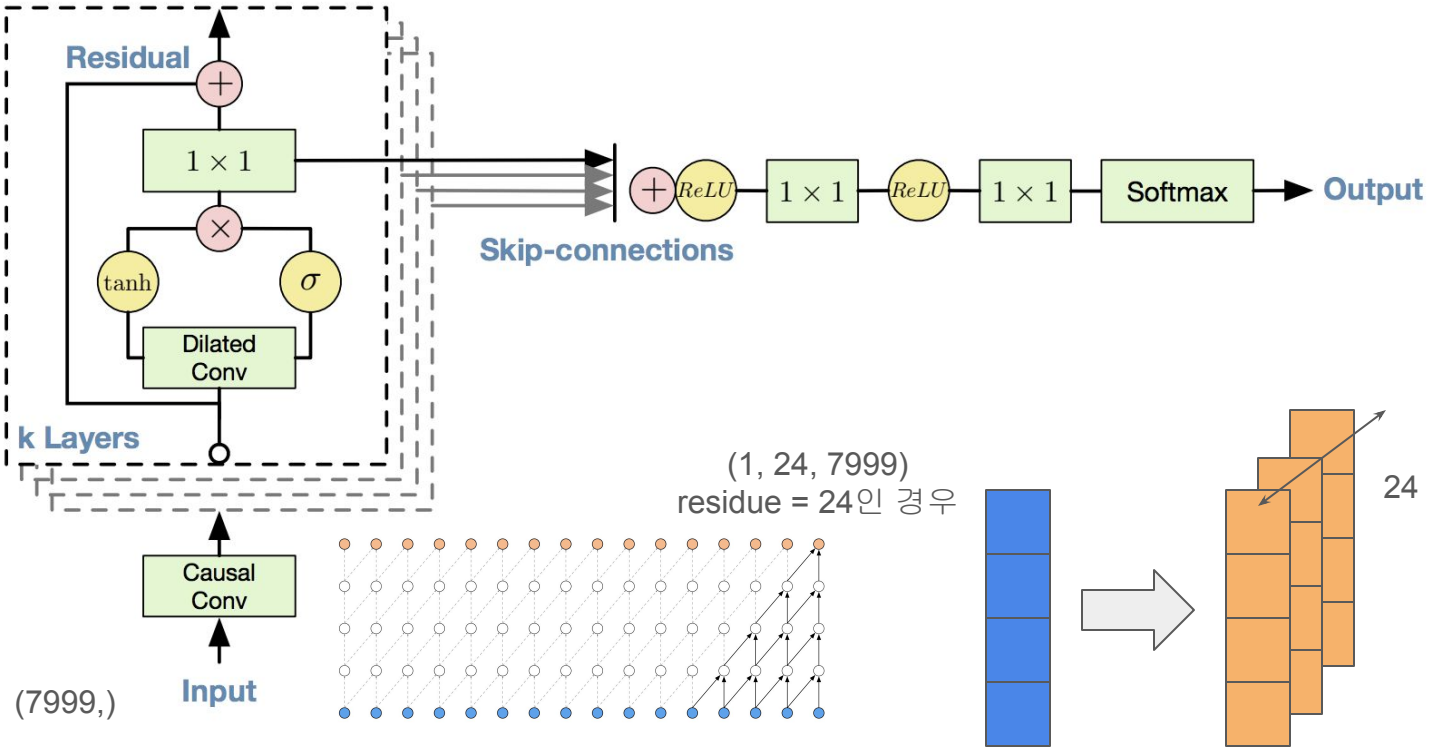
https://arxiv.org/pdf/1609.03499.pdf

Residual and Skip connection



https://arxiv.org/pdf/1609.03499.pdf

Residual and Skip connection



https://arxiv.org/pdf/1609.03499.pdf

Residual and Skip connection

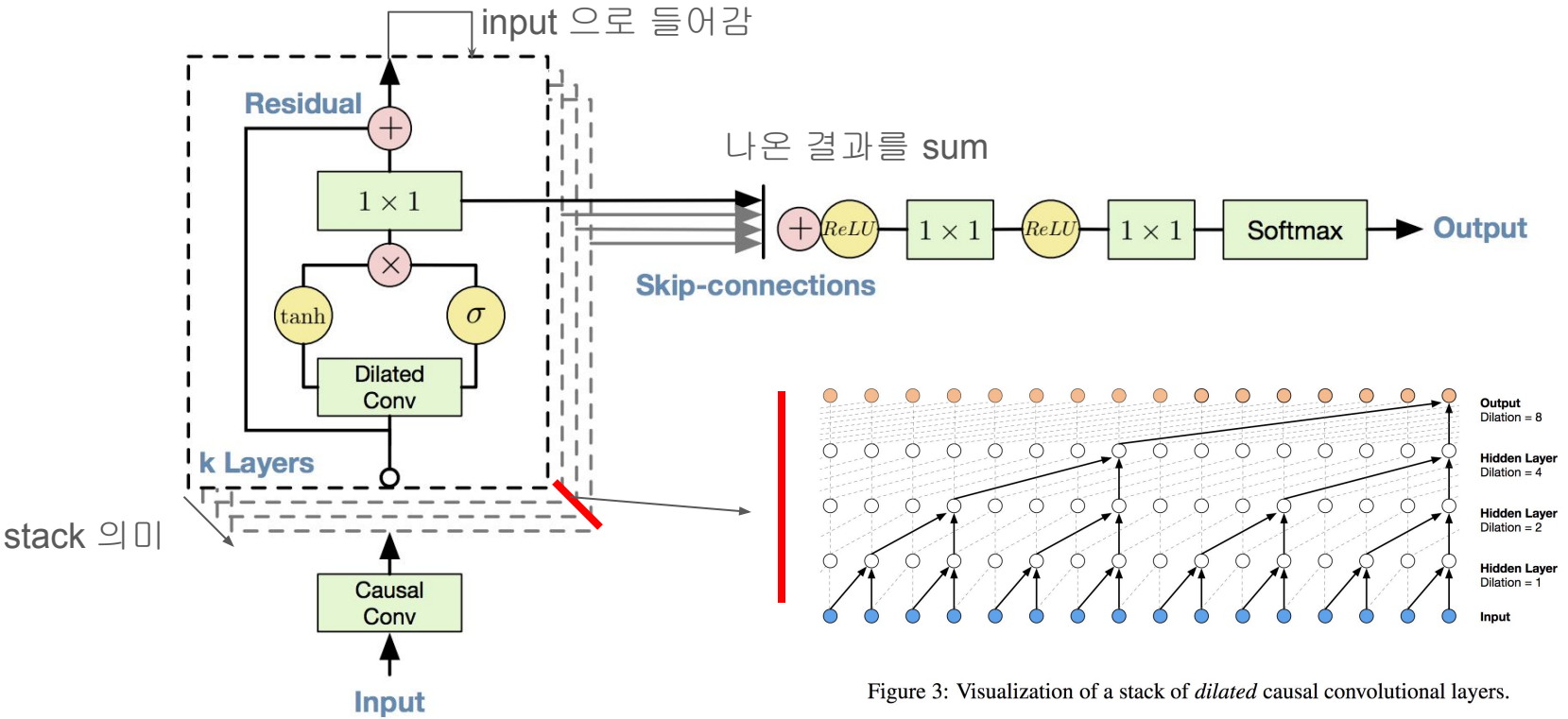
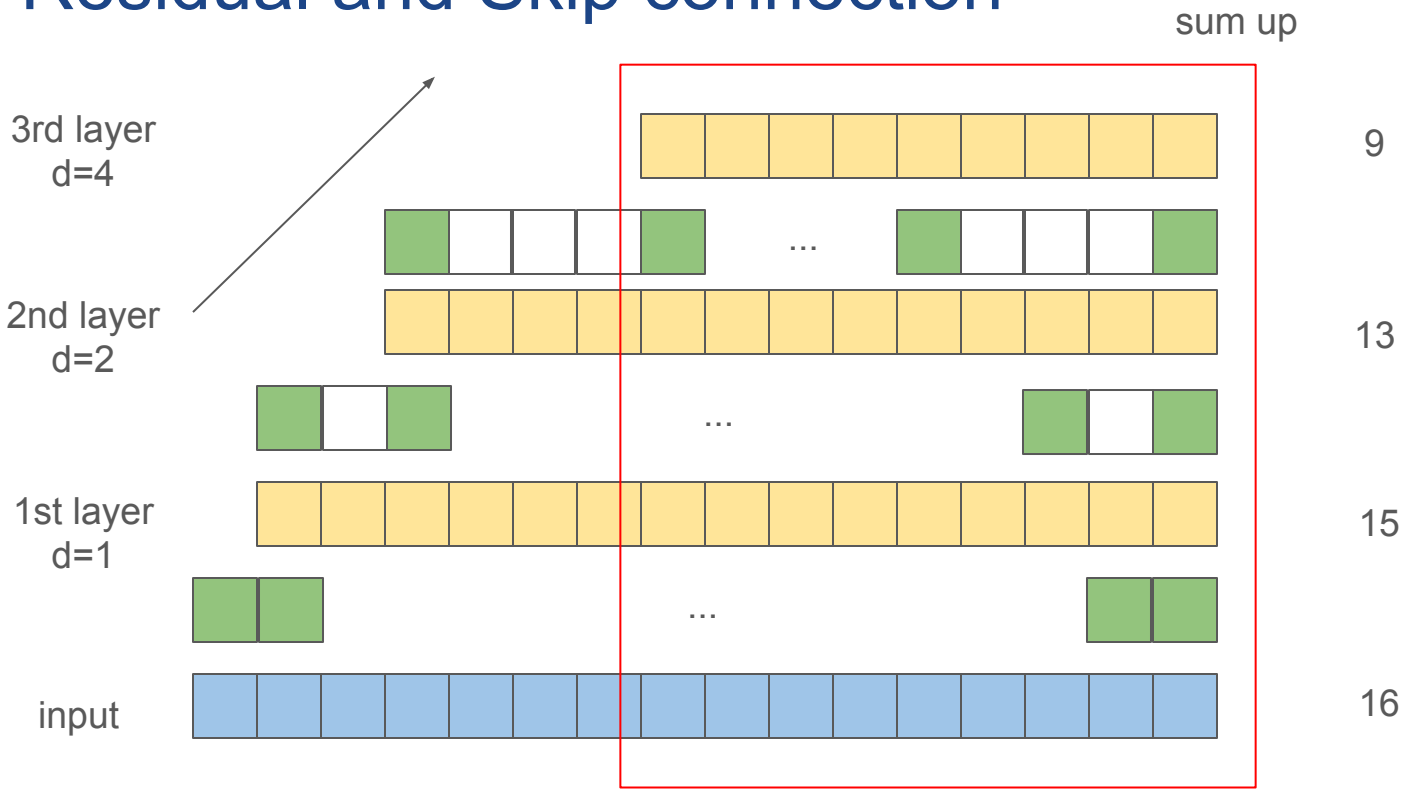


Figure 3: Visualization of a stack of *dilated* causal convolutional layers.

https://arxiv.org/pdf/1609.03499.pdf

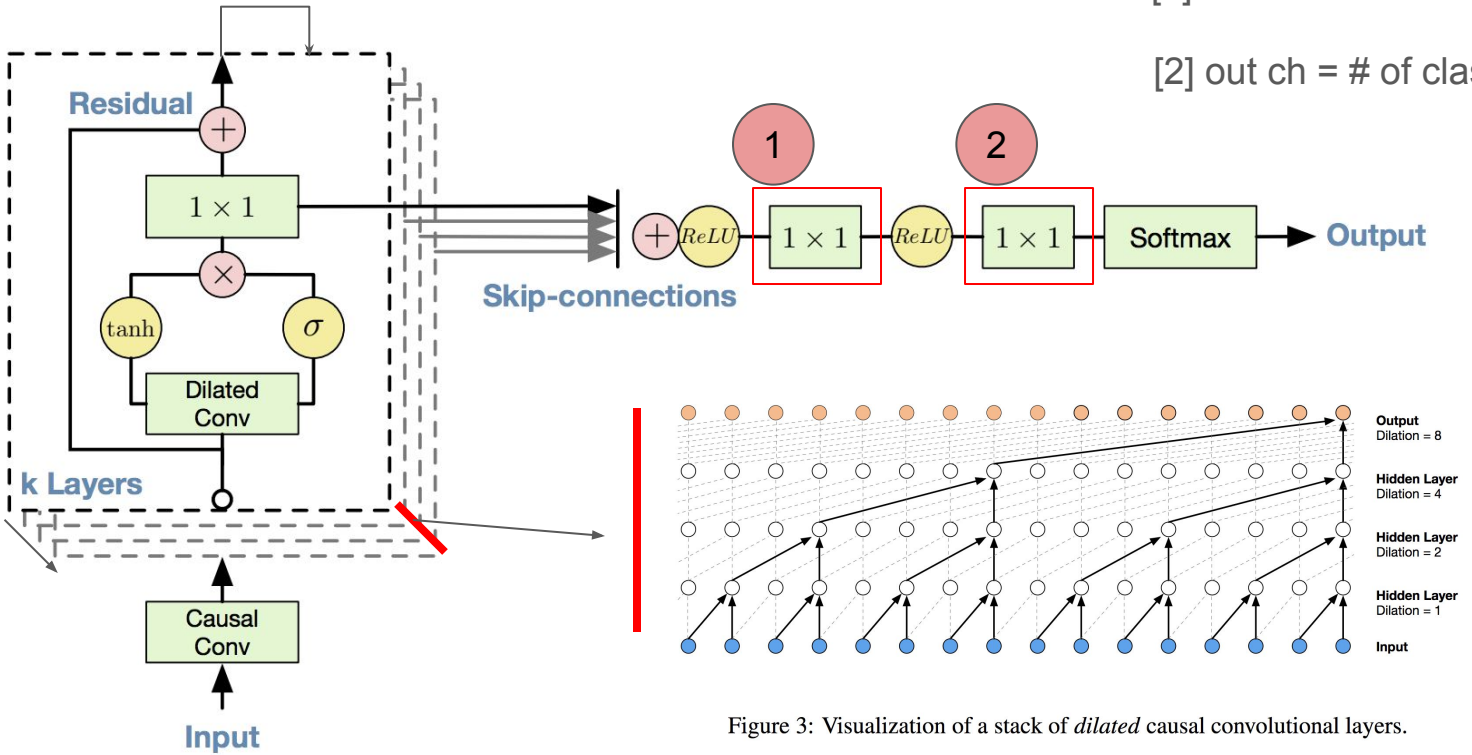
Residual and Skip connection



Residual and Skip connection

[1] in ch = out ch

[2] out ch = # of classes



stack 의 미

Figure 3: Visualization of a stack of *dilated* causal convolutional layers.

https://arxiv.org/pdf/1609.03499.pdf

code를 통한 WaveNet 이해 및 구현

- jupyter code 참조

END OF DOCUMENT