Sequence to sequence models

Basic models

deeplearning.ai

Sequence to sequence model

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ $x^{<4>}$ $x^{<5>}$

Jane visite l'Afrique en septembre

$\longrightarrow$ Jane is visiting Africa in September.

$y^{<1>}$ $y^{<2>}$ $y^{<3>}$ $y^{<4>}$ $y^{<5>}$ $y^{<6>}$

$a^{<0>}$

$x^{<1>}$

$x^{<T_x>}$
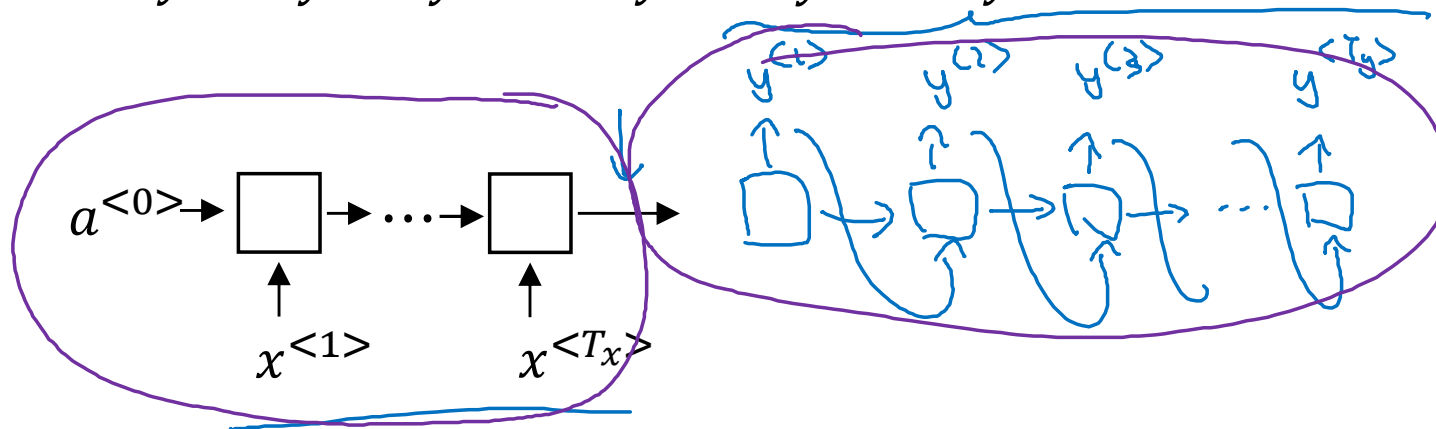
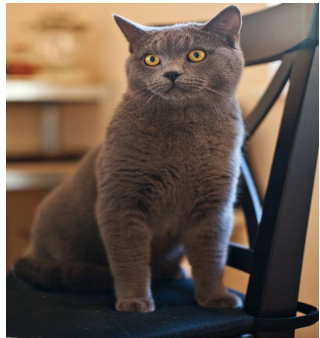$\hat{y}^{<1>}$ $\hat{y}^{<2>}$ $\hat{y}^{<3>}$ $\hat{y}^{<T_y>}$

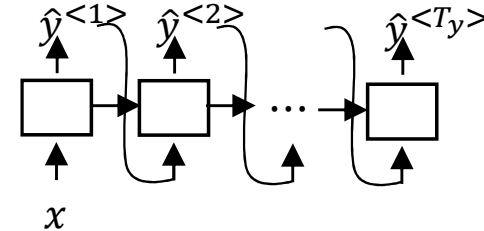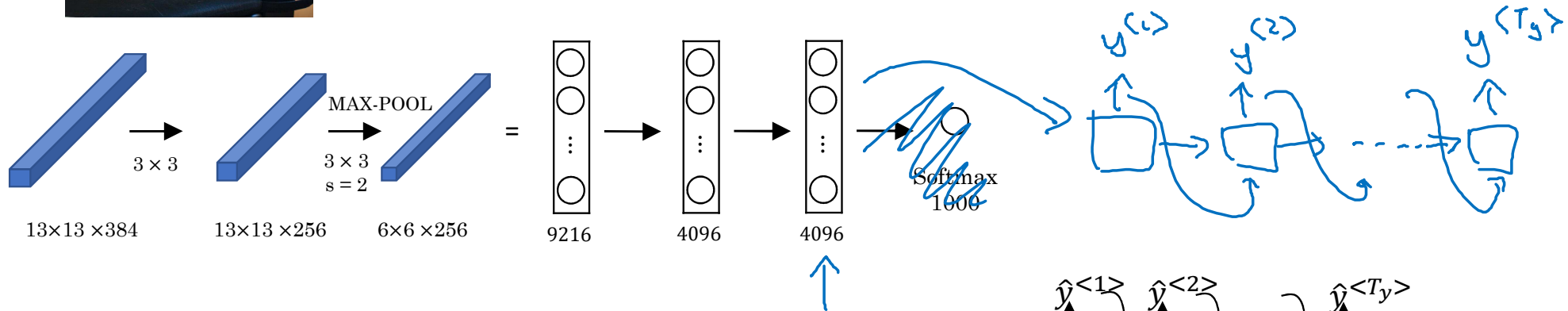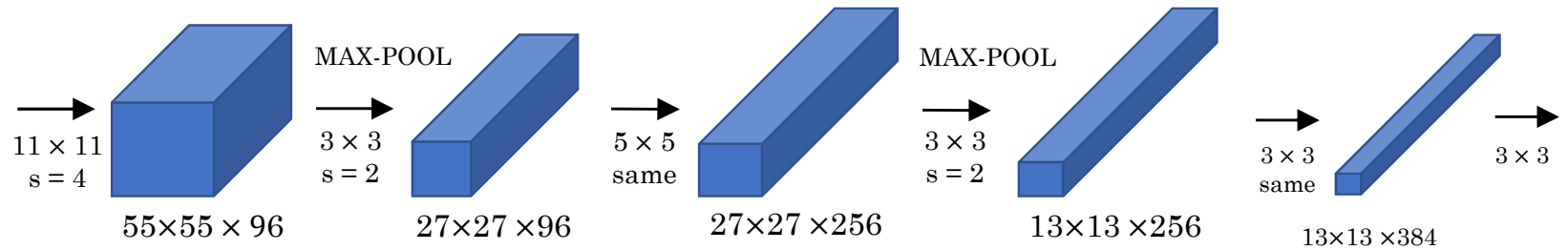[Sutskever et al., 2014. Sequence to sequence learning with neural networks]

[Cho et al., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation]

Andrew Ng

Image captioning

$y^{<1>}\ y^{<2>}\quad y^{<3>}\quad y^{<4>}\quad y^{<5>}\quad y^{<6>}$

A   cat   sitting   on   a   chair

MAX-POOL

11×11
$s = 4$

3×3
$s = 2$

5×5
same

MAX-POOL

3×3
$s = 2$

3×3
same

3×3

$55{\times}55 \times 96$ $27{\times}27 \times 96$ $27{\times}27 \times 256$ $13{\times}13 \times 256$ $13{\times}13 \times 384$

3×3 MAX-POOL 3×3 $s = 2$ $=$

$y^{(1)}$ $y^{(2)}$ $y^{<T_y>}$

$13{\times}13 \times 384$ $13{\times}13 \times 256$ $6{\times}6 \times 256$ 9216 4096 4096

Softmax
1000

$\hat{y}^{<1>}\quad \hat{y}^{<2>}\quad \hat{y}^{<T_y>}$

x

[Mao et. al., 2014. Deep captioning with multimodal recurrent neural networks]
[Vinyals et. al., 2014. Show and tell: Neural image caption generator]
[Karpathy and Li, 2015. Deep visual-semantic alignments for generating image descriptions]

Andrew Ng