Neural Network

- A network connecting numerous neurons
Neural Network

- A network connecting numerous neurons
Analogy

- Imagine a neural network as a map
Analogy

- Imagine a neural network as a map
- Imagine a neuron as a place
Analogy

- Imagine a neural network as a map
- Imagine a neuron as a place
- Imagine yourself as the information flow
Analogy

- Imagine a neural network as a map
- Imagine a neuron as a place
- Imagine yourself as the information flow
Suppose you (information flow) wants to reach Bakery (neuron B) from City Hall (neuron A), what will you do?
Suppose you (information flow) wants to reach Bakery (neuron B) from City Hall (neuron A), what will you do?

You have to follow the path of network!
Analogy

- Suppose you (information flow) wants to reach Bakery (neuron B) from City Hall (neuron A), what will you do?
- You have to follow the path of network!
- What if there is a highway connecting Bakery and City Hall directly?

Not drawn to scale
Highway Networks

Allowing direct pass (highway) between neurons in different layers.
Highway Networks

Allowing direct pass (highway) between neurons in different layers.
Highway Networks

Original network:

$$z_1 = \sigma \left( \sum_{n=1}^{N_1} w_{n}^{1} x_{n} + b \right) \tag{1}$$
Highway Networks

Original network:

\[ z_1 = \sigma \left( \sum_{n=1}^{\infty} w_n^1 x_n + b \right) \tag{1} \]

Highway network:

\[ z_1 = T \sigma \left( \sum_{n=1}^{\infty} w_n^1 x_n + b \right) + (1 - T)x_1 \tag{2} \]
Highway Networks

Original network:

\[ z_1 = \sigma \left( \sum_{n=1} w_n^1 x_n + b \right) \] (1)

Highway network:

\[ z_1 = T \sigma \left( \sum_{n=1} w_n^1 x_n + b \right) + (1 - T)x_1 \] (2)

Gating function:

\[ T = \sigma \left( \sum_{n=1} w_n' x_n + b' \right) \] (3)
Highway Networks

- Remember the shape of sigmoid function.
Highway Networks

- Remember the shape of sigmoid function.

\[ \text{sig}(t) = \frac{1}{1+e^{-t}} \]
Highway Networks

- Remember the shape of sigmoid function.

\[ \text{sig}(t) = \frac{1}{1+e^{-t}} \]

- We can set bias \( b' \) to negative values such that gating value \( T \to 0 \).
Benefits of Highway networks

- Enable training of very deep neural networks (e.g., hundreds of layers)

Benefits of Highway networks

- Enable training of very deep neural networks (e.g., hundreds of layers)

Motivation: Does depth matter for deep learning?
Residual Networks

Motivation: Does depth matter for deep learning?

Residual Networks

- Motivation: Does depth matter for deep learning?

We need new architecture to make depth matter.
Residual Networks

- We need new architecture to make depth matter.
- Suppose you have a plain 2-layer network $\mathcal{H}$.

```
\[ H(x) = \text{relu} \left( \text{weight layer} \left( \text{relu} \left( \text{weight layer}(x) \right) \right) \right) \]
```

Residual Networks

- We need new architecture to make depth matter.
- Suppose you have a plain 2-layer network $\mathcal{H}$.
- We use a new building block which forces the previous 2-layer $\mathcal{F}$ to learn the residual $\mathcal{H} - x$. 

Residual Networks

- We need new architecture to make depth matter.
- Suppose you have a plain 2-layer network $\mathcal{H}$.
- We use a new building block which forces the previous 2-layer $\mathcal{F}$ to learn the residual $\mathcal{H} - x$.

\[ H(x) = F(x) + x \]


Renjie Liao (UofT)
What we have done?
Residual Networks

- What we have done?

\[ x \rightarrow \text{weight layer} \rightarrow \text{relu} \rightarrow \text{weight layer} \rightarrow \text{relu} \rightarrow H(x) \]
Residual Networks

What we have done?

\[
H(x) = F(x) + x
\]

Based on this building block, we can do some crazy things like...
Based on this building block, we can do some crazy things like...

Revolution of Depth

- AlexNet, 8 layers (ILSVRC 2012)
- VGG, 19 layers (ILSVRC 2014)
- ResNet, 152 layers (ILSVRC 2015)

PASCAL VOC Challenge Results

Revolution of Depth

Engines of visual recognition


<table>
<thead>
<tr>
<th>task</th>
<th>2nd-place winner</th>
<th>MSRA</th>
<th>margin (relative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ImageNet Localization (top-5 error)</td>
<td>12.0</td>
<td>9.0</td>
<td>27%</td>
</tr>
<tr>
<td>ImageNet Detection (mAP@.5)</td>
<td>53.6 <strong>absolute 8.5% better!</strong></td>
<td>62.1</td>
<td>16%</td>
</tr>
<tr>
<td>COCO Detection (mAP@.5:.95)</td>
<td>33.5</td>
<td>37.3</td>
<td>11%</td>
</tr>
<tr>
<td>COCO Segmentation (mAP@.5:.95)</td>
<td>25.1</td>
<td>28.2</td>
<td>12%</td>
</tr>
</tbody>
</table>
Thanks!