Highway Networks and Residual Networks

Renjie Liao

University of Toronto

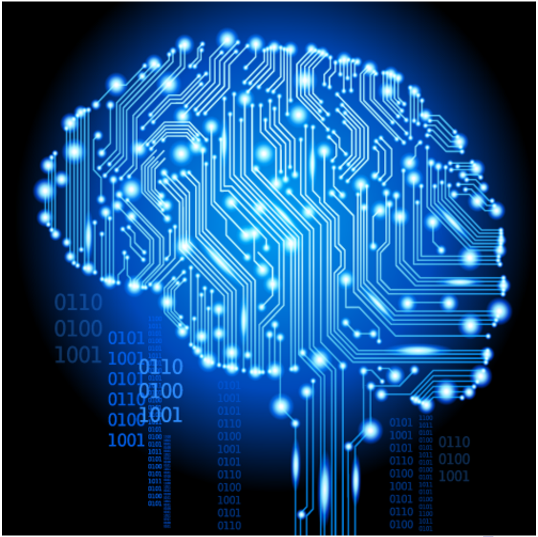
Jan 26, 2016

Neural Network

- A network connecting numerous neurons

Neural Network

- A network connecting numerous neurons



Analogy

- Imagine a neural network as a map

Analogy

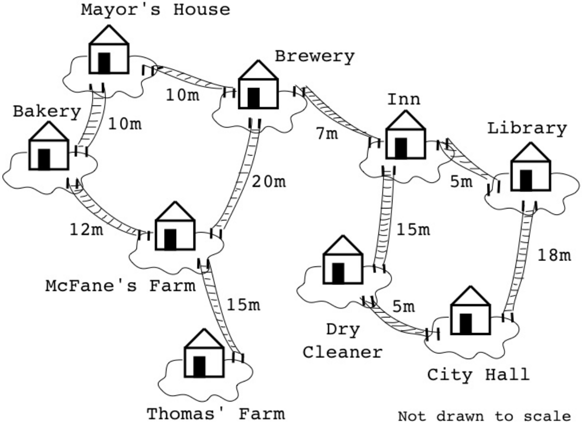
- Imagine a neural network as a map
- Imagine a neuron as a place

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



Analogy

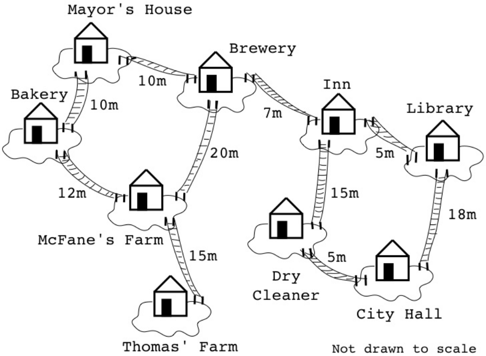
- Suppose you (information flow) wants to reach Bakery (neuron B) from City Hall (neuron A), what will you do?

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!

Analogy

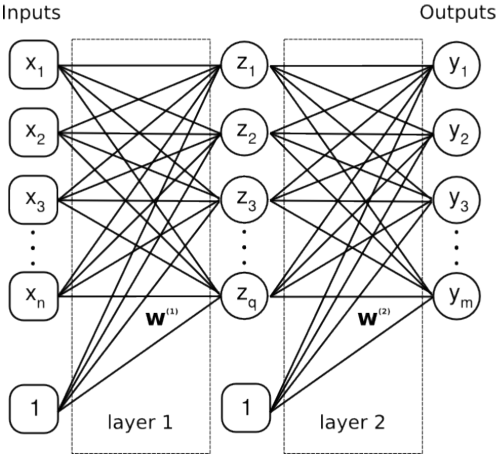
- Suppose you (information flow) wants to reach Bakery (neuron B) from City Hall (neuron A), what will you do?
- You have to follow the path of network!
- What if there is a highway connecting Bakery and City Hall directly?



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

Highway Networks

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



Highway Networks

Original network:

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

Highway Networks

Original network:

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

Highway network:

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

Highway Networks

Original network:

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

Highway network:

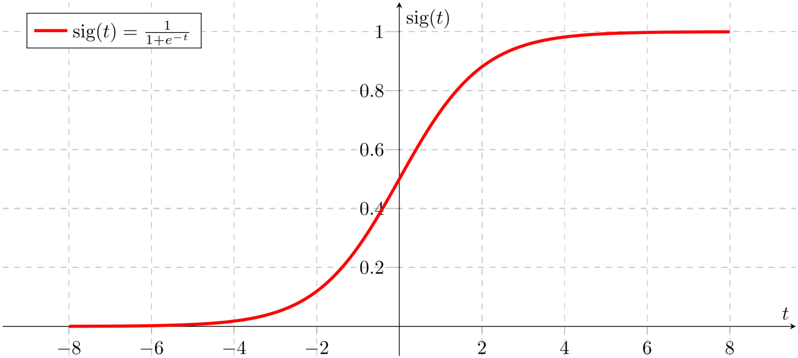
$$z_1 = T\sigma\left(\sum_{n=1} w_n^1 x_n + b\right) + (1 - T)x_1 \tag{2}$$

Gating function:

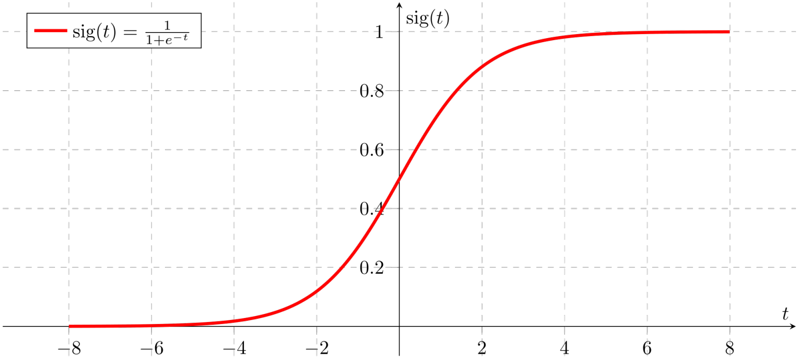
$$T = \sigma\left(\sum_{n=1} w_n' x_n + b'\right) \tag{3}$$

- Remember the shape of sigmoid function.

- Remember the shape of sigmoid function.



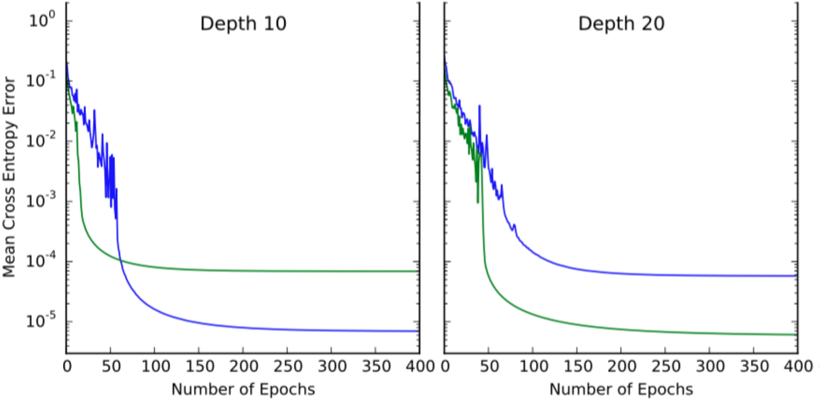
- Remember the shape of sigmoid function.



- We can set bias b' to negative values such that gating value $T \to 0$.

Benefits of Highway networks

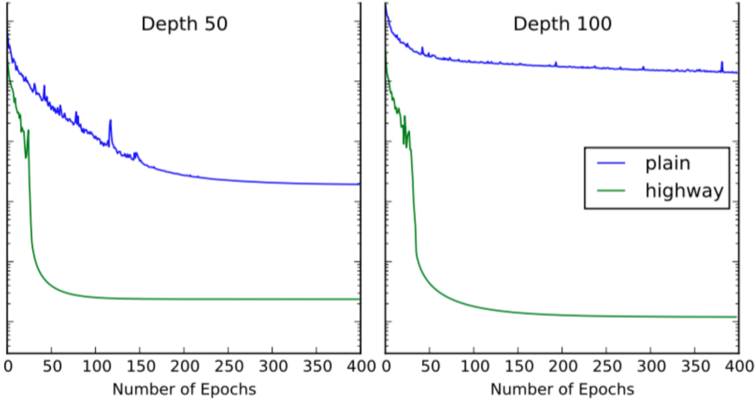
- Enable training of very deep neural networks (e.g., hundreds of layers)



"Srivastava, R.K., Greff, K. and Schmidhuber, J., 2015. Highway Networks. arXiv preprint arXiv:1505.00387".

Benefits of Highway networks

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

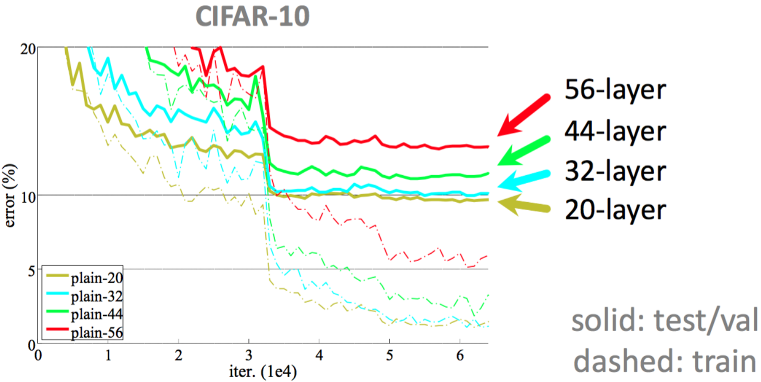


"Srivastava, R.K., Greff, K. and Schmidhuber, J., 2015. Highway Networks. arXiv preprint arXiv:1505.00387".

Residual Networks

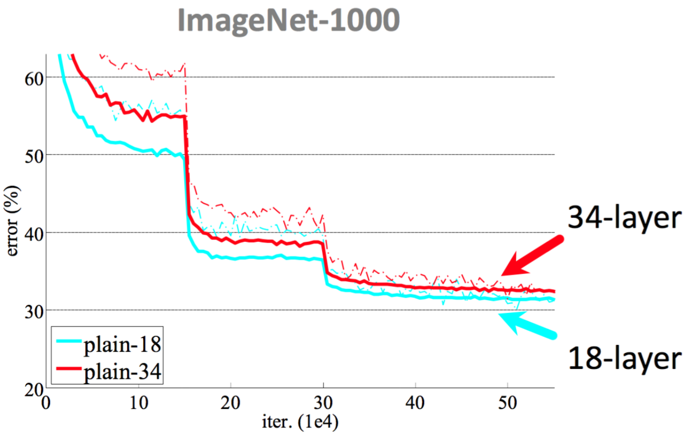
- Motivation: Does depth matter for deep learning?

- Motivation: Does depth matter for deep learning?



"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

- Motivation: Does depth matter for deep learning?



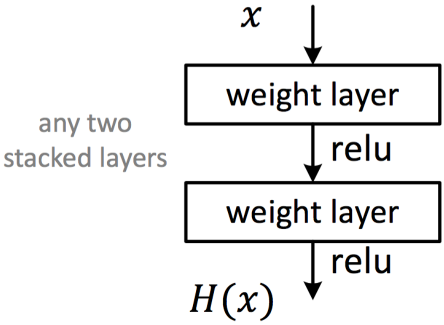
"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

Residual Networks

- 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 $\mathcal{H}$.



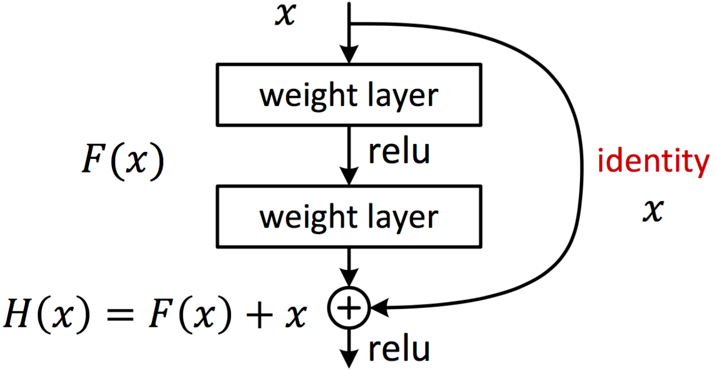
"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

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 $\mathcal{H} - x$.



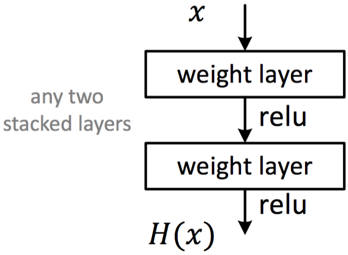
"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

Residual Networks

- What we have done?

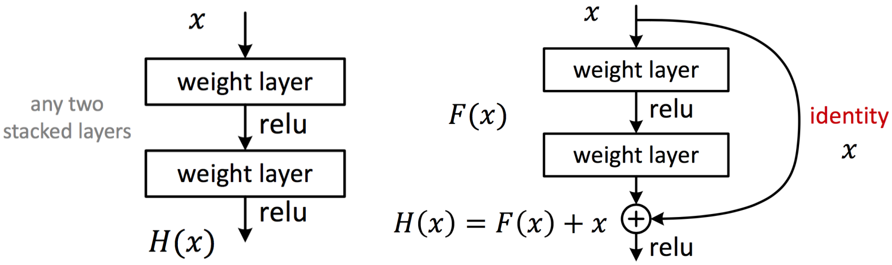
Residual Networks

- What we have done?



Residual Networks

- What we have done?

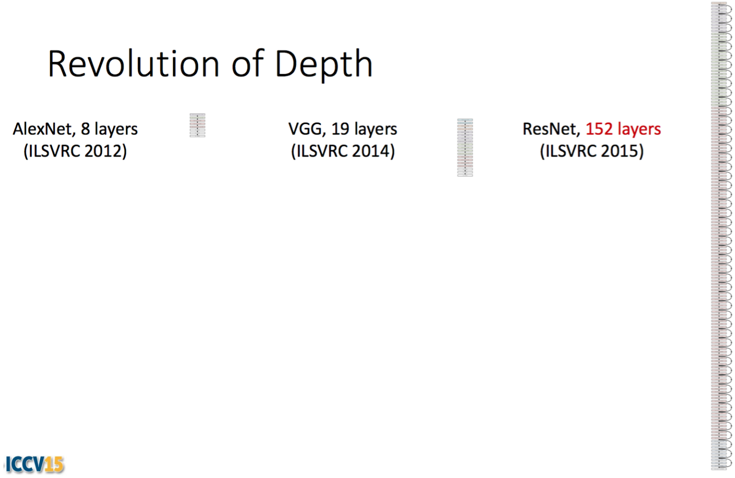


"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

Residual Networks

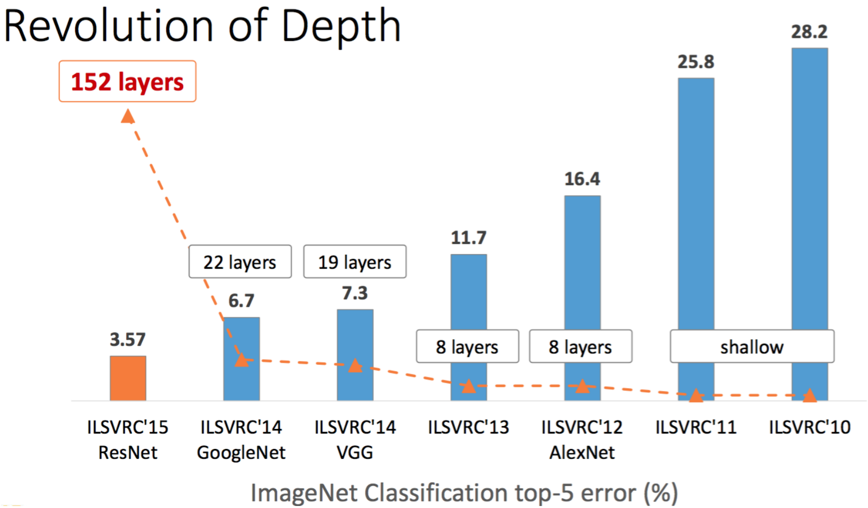
- Based on this building block, we can do some crazy things like...

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



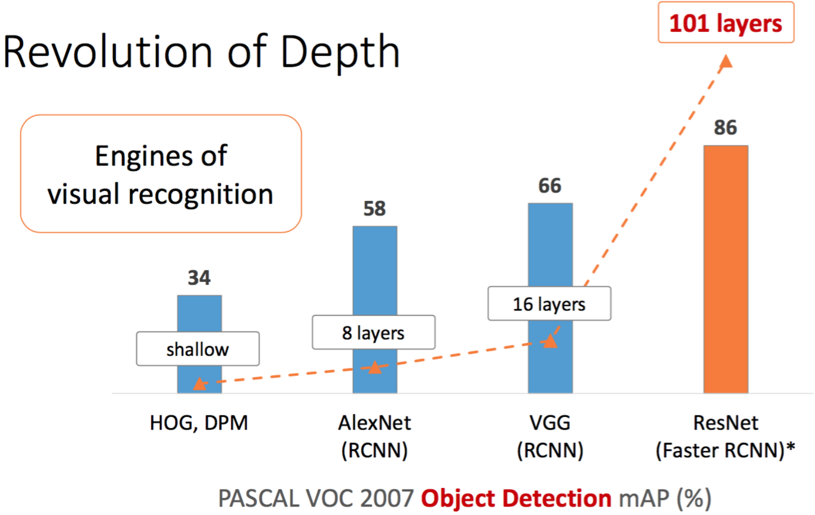
"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

Image Net Challenge Results



"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

PASCAL VOC Challenge Results



"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

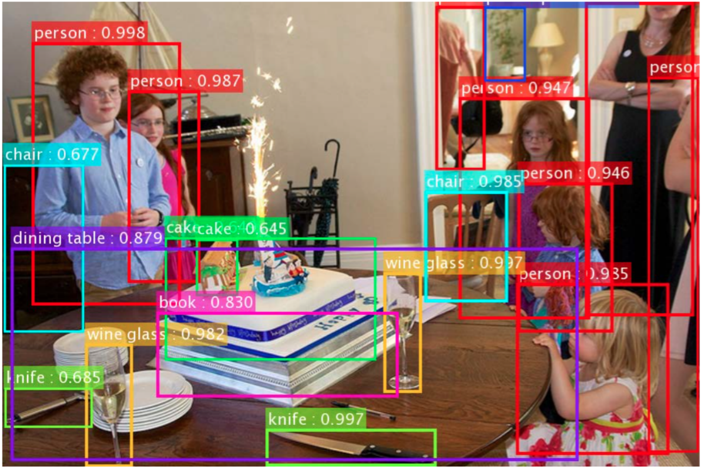
More Results

task	2nd-place winner	MSRA	margin (relative)
ImageNet Localization (top-5 error)	12.0	9.0	27%
ImageNet Detection (mAP@.5)	53.6	62.1	16%
COCO Detection (mAP@.5:.95)	33.5	37.3	11%
COCO Segmentation (mAP@.5:.95)	25.1	28.2	12%

absolute 8.5% better!

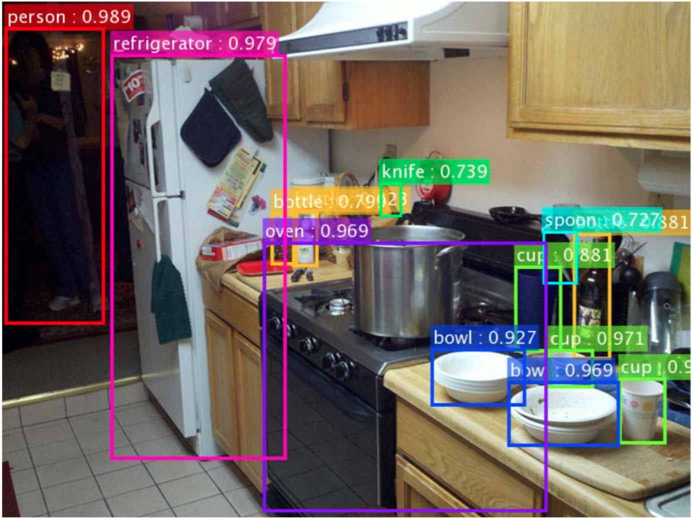
"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

More Results



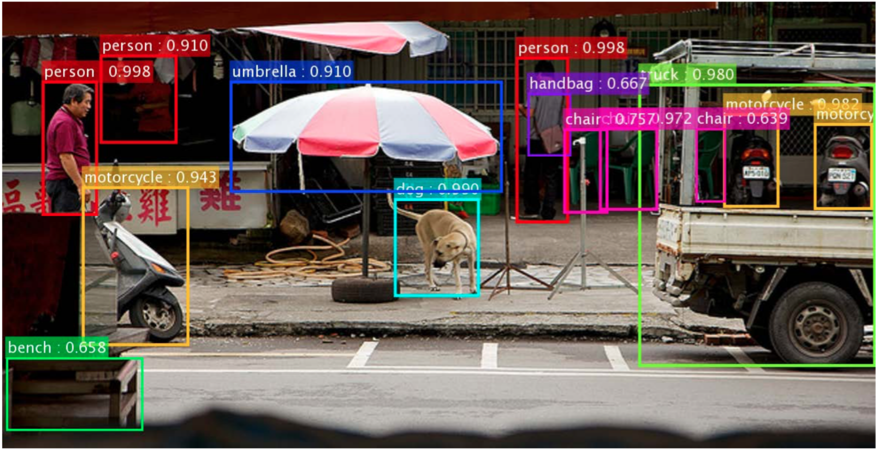
"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

More Results



"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

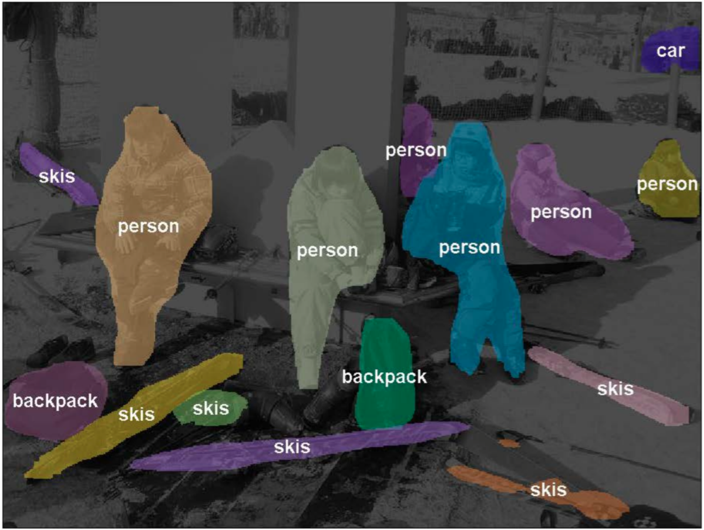
More Results



"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

More Results

input



"He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385".

Thanks!