Capsule Networks

Based on NIPS 2017 Paper

Dynamic Routing Between Capsules

By Sara Sabour, Nicholas Frosst, Geoffrey E Hinton

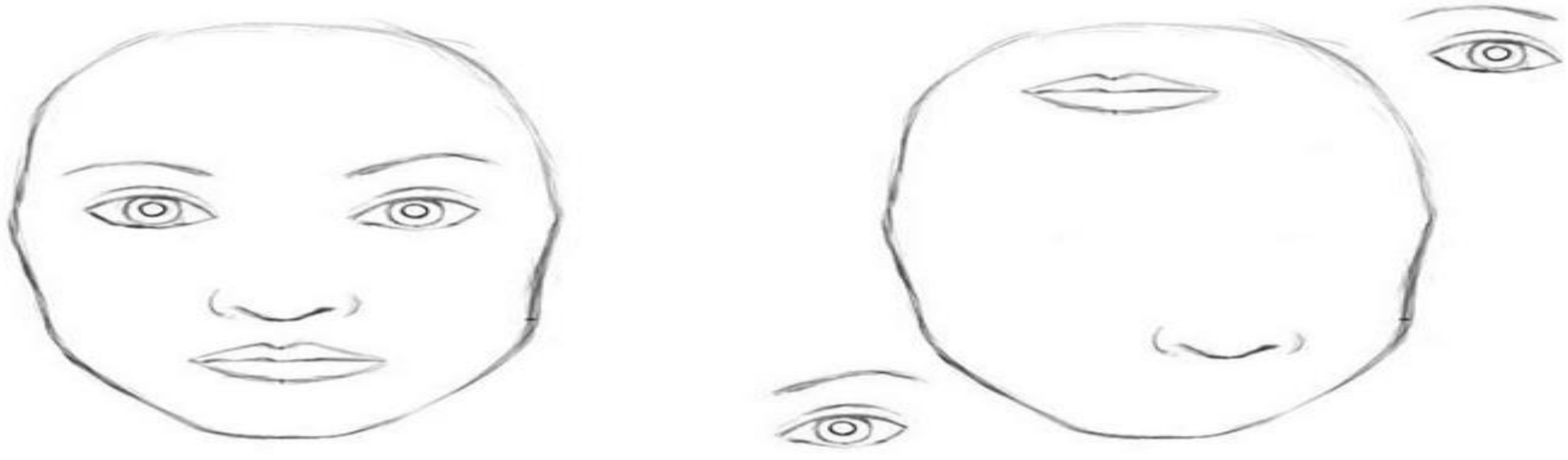
October 2017: *https://arxiv.org/abs/1710.09829*

Sreya Francis

What is wrong with Convolutional Neural Networks?

- If the images have rotation, tilt or any other different **orientation**, then CNNs have poor performance.

- In CNN, each layer understands an image at a much more granular level(Slow increase in **receptive field**)



To a CNN, both pictures are similar, since they both contain similar elements.

- **Pooling Layer** used in CNN creates positional invariance which triggers false positive for images.

- CNNs are easily **fooled**.

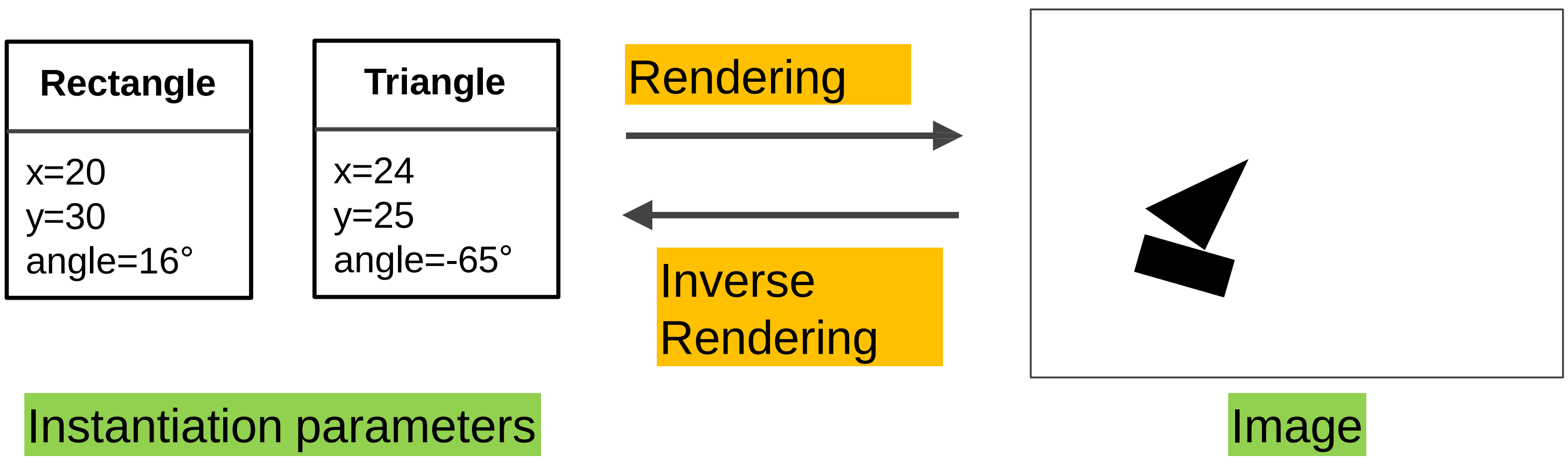
- In CNN, the internal data representation do not take into account **important spatial hierarchies** between simple and complex objects.

How to solve this problem?

- **Inverse Graphics Approach** : Hard code 3D world into a neural net.

•**Rendering**: Computer Graphics takes internal hierarchical representation of objects and produces an image.

•**Inverse Rendering**: Human Brain does the opposite. Capsule networks follow a similar approach.



CNN	Capsule Network
Scalar output feature detector	Vector output capsule
Max pooling	Routing by agreement
Aims for Invariance	Aims for equivariance

Algorithm Used: Dynamic Routing B/w Capsules

•**Capsules**: New building block in DL to better model hierarchical relationships.

•**Dynamic Routing**: Allows capsules to communicate with each other and create scene graphs similar to scene graphs in computer graphics.

•Capusle network does not depend on view angle.

•The capsule network is much better than other models at telling that images in top and bottom rows belong to the same classes, only the view angle is different.

•The latest papers decreased the error rate by a whopping 45%.

•Source:https://openreview.net/pdf?id=HJWLfGWRb

Pros:

- Reaches high accuracy on MNIST, and promising on CIFAR10
- Requires less training data
- Position and pose information are preserved (equivariance)
- This is promising for image segmentation and object detection
- Routing by agreement is great for overlapping objects (explaining away)
- Capsule activations nicely map the hierarchy of parts
- Offers robustness to affine transformations
- Activation vectors are easier to interpret (rotation, thickness, skew…)

Cons:

- Not state of the art performance on CIFAR10 (but it’s a good start)
- Not tested yet on larger images (e.g., ImageNet): will it work well?
- Slow training, due to the inner loop (in the routing by agreement algorithm).
- Data flow is more complicated.
- Harder to calculate gradients.
- Model may suffer more from vanishing gradients.
- A Caps-Net cannot see two very close identical objects
    - This is called “crowding”, and it has been observed as well in human vision

Implementation:

- Keras with Tensorflow as backend :https://github.com/XifengGuo/CapsNet-Keras
- TensorFlow: https://github.com/naturomics/CapsNet-Tensorflow
- PyTorch: https://github.com/gram-ai/capsule-networks