

# ImageNet Paper Report:

## Dataset:

- Contains 15 million images from 2200 categories
- Variable resolution image converted to 256 \* 256
- Model is trained on this dataset **centered raw RGB values**

## Architecture:

- Neural network architecture has **60 million parameters** consisting of **8 layers with 5 convolutional and 3 fully-connected**. The output of the last fully-connected layer is fed to a 1000-way softmax
- **ReLU**
  - Reached training error 25% 6 times faster than tanh activation function.
- **Multiple GPUs** (2 GTX 580 GPU 3GB)
  - The kernels of the second, fourth, and fifth convolutional layers are connected only to those kernel maps in the previous layer which reside on the same GPU. The kernels of the third convolutional layer are connected to all kernel maps in the second layer. The neurons in the fully-connected layers are connected to all neurons in the previous layer
  - The **columnar dependent architecture** is built on this which helped reduce the error
- **Local Response Normalization**
  - Despite ReLU not needing normalization, applying this helps the model generalize better
- **Overlapping Pooling**
  - This occurs when kernel size > stride and this pooling finds hard to overfit

## Reduce Overfitting:

- **Data Augmentation**
  - Generating image translations and horizontal reflections
  - Altering the intensities of the RGB channels in training images using **PCA**
- **Dropout**
  - Setting to zero the output of each hidden neuron with probability 0.5
  - Forced to learn more robust features

## Details of learning:

- Small amount of weight decay was important for the model to learn (**Momentum SGD**)
- Weights initialized using zero mean gaussian distribution with SGD = 0.01 and biases were set to constant 1 and 0 in some places

## Results:

- Results on ILSVRC-2010 are: Our network achieves top-1 and top-5 test set error rates of 37.5% and 17.0%.
- Results on ILSVRC-2012 is top-5 testing error of 15.3%.