ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky, Ilya Sutskever & Geoffrey E. Hinton

NIPS (2012)

Presenter: Ogenna Esimai

CSE 6369 - Special Topics in Advanced Intelligent Systems, Human Computer Interactions

October 21, 2021

© 2021 Ogenna Esimai

Outline

- Introduction to the Problem
- Background
- Novelty and Innovation
- Methodology and Relation to HCI
- Applications
- Conclusion and Critical Thoughts

© 2021 Ogenna Esimai

Introduction to the Problem

- Competing in contest ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC-2012)
- Speed and lower error rates
- Won
- Published a variant of their strategy

© 2021 Ogenna Esimai

Background

- Object recognition contest
  - ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC-2012)
  - annual
  - started in 2010
- ImageNet dataset
  - over 15 million labeled high-resolution images
  - ~ 22,000 categories

© 2021 Ogenna Esimai

Novelty and Innovation

- A much better way of object recognition
- Deep Convolutional Neural Network constructed in novel way
- Trained one of the largest networks to date
  - subsets of ImageNet in ILSVRC-2010 and ILSVRC-2012
- Implemented 2D convolution, highly-optimized, GPU
  - Publicly available
- Novel or unusual features of architecture
  - ReLU nonlinearity
  - training on multiple GPUs
  - local response normalization
  - overlapping pooling

© 2021 Ogenna Esimai

Methodology and Relation to HCI

- Convolutional Neural Network
- Special features of architecture
  - ReLU nonlinearity: f(x) = max(0, x)
  - training on multiple GPUs
  - local response normalization
  - overlapping pooling
- Reducing overfitting
  - data augmentation
    - image translations and horizontal reflections
    - altering the intensities of channels training images
  - dropout

© 2021 Ogenna Esimai

Methodology and Relation to HCI (2)

- Architecture
  - eight layers with weights
  - first five are convolutional
  - then three are fully connected
  - output of last fully-connected layer feeds to a 1000-way softmax
  - produces a distribution over the 1000 class labels

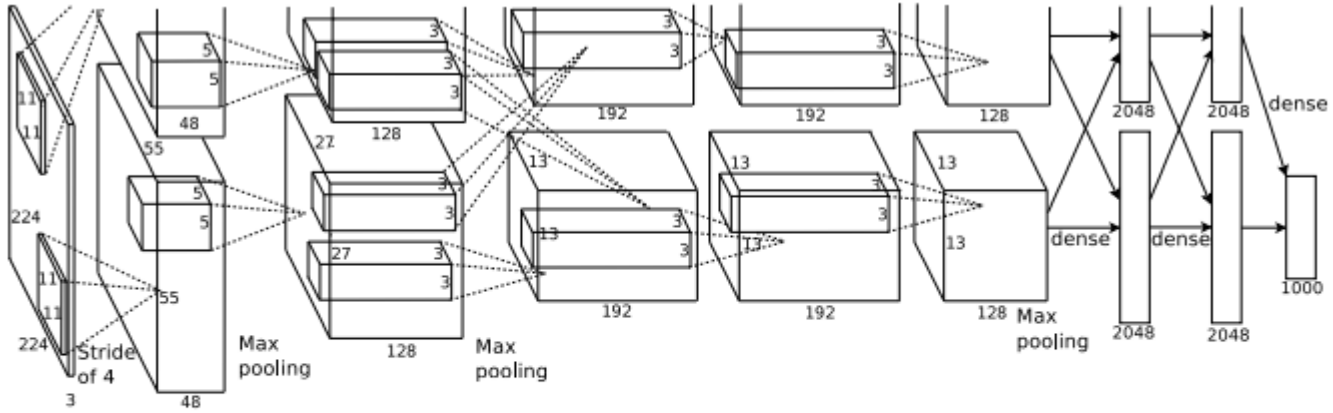
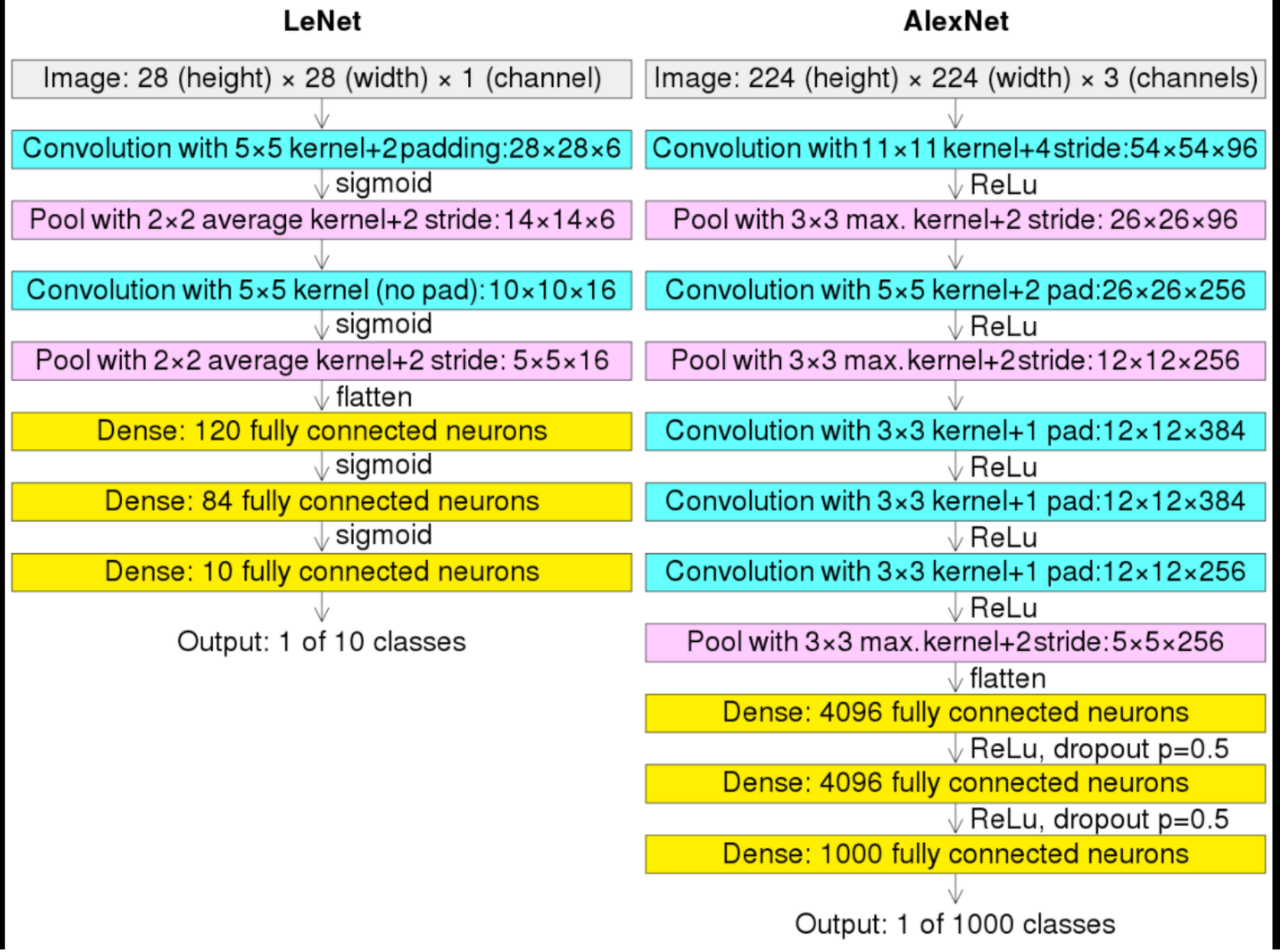


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

(Accessibility) Figure – A schematic in black and white of the architecture of the Convolutional Neural Network implemented in the paper. From left to right, after a common input of images, the network runs along a top and a bottom pipeline. These 2 pipelines are run by two different Graphical Processing Units (GPUs), are mostly independent, and communicate only at certain points in the architecture. **Figure on next slide** – right hand side of picture shows a color rendering of the different layers and architecture of the same Convolutional Neural Network. Gray – input layer, Blue – convolutional layer, Pink – pooling layer, Yellow – dense, fully connected layer.

© 2021 Ogenna Esimai



Comparison of the LeNet and AlexNet convolution, pooling and dense layers

More details

Cmglee - Own work

CC BY-SA 4.0

Courtesy of CMG Lee.
https://en.wikipedia.org/wiki/Convolutional_neural_network#/media/File:Comparison_image_neural_networks.svg

© 2021 Ogenna Esimai

Methodology and Relation to HCI (3)

- Reduced error rates
- Relation to HCI
  - far-reaching
  - object recognition in AR/VR

Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

Table 1: Comparison of results on ILSVRC-2010 test set. In *italics* are best results achieved by others.

Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
SIFT + FVs [7]	—	—	*26.2%*
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	**16.4%**
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	**15.3%**

Table 2: Comparison of error rates on ILSVRC-2012 validation and test sets. In *italics* are best results achieved by others. Models with an asterisk* were "pre-trained" to classify the entire ImageNet 2011 Fall release. See Section 6 for details.

© 2021 Ogenna Esimai

Applications

- Application area
  - Object recognition in CV and related, broad

Proceeding of Fifth International Conference on Microelectronics, Computing and Communication Systems pp 77-89 | Cite as

AlexNet-CNN Based Feature Extraction and Classification of Multiclass ASL Hand Gestures

Authors | Authors and affiliations

Abul Abbas Barbhuiya, Ram Kumar Karsh, Samiran Dutta

Conference paper
First Online: 10 September 2021

82 Downloads

© 2021 Ogenna Esimai

Conclusion and Critical Thoughts

- Conclusion
  - Improvement over state-of-the-art at the time
- Future Works
  - Highly impactful
  - AlexNet
- Critical Thoughts
  - Solid contributions to HCI through object recognition and CV
  - Extensive use of jargon likely from intended audience
  - Provides example of how existing techniques + innovation -> groundbreaking results

© 2021 Ogenna Esimai