Chapter 5: ConvNets and ImageNet

1. He combined ideas from Fukushima's neocognitron with the back-propagation algorithm to create the semi-eponymous "LeNet" – one of the earliest ConvNets. Who was he? Who was his postdoctoral advisor? And, in addition to developing "LeNet", what was his contribution to the field of neural network research?

Yann Lecun created "LeNet." Lecun's postdoctoral advisor is Geoffrey Hinton (81). Lecun worked on a system that recognizes handwritten digits. It was quickly adapted by the banking industry and the Postal Service. Lecun's continued work on ConvNets, despite declining popularity and support in the field, paved the way for the public's renewed interest in neural networks (81-82).

2. What is WordNet? Please answer by referencing its creator (the human most responsible for its existence), saying something about its use, and presenting, in just two or three sentences, a high-level description of its structure?

Fei-Fei Li, a Princeton computer science professor, worked on developing WordNet with Princeton psychology professor George Miller due to disappointment with the visually narrowed programs being entered in the PASCAL image recognition competitions. WordNet is a hierarchical word database that records the relationships between words from "most specific to most general." It is useful in natural language processing (83-84).

3. What is ImageNet? Please answer by referencing its creator (the human most responsible for its existence), saying something about its use, and presenting a high-level description of its structure, being sure to mention WordNet in doing so.

ImageNet is a program used for object recognition that was created by Fei-Fei Li. The inspiration for the program came from WordNet. However, instead of linking words together, ImageNet links images to nouns, matching a large sample of images to a singular noun (84).

4. Specifically, what role did ImageNet play in the advancement of computer programs dedicated to the problem of "object recognition."

ImageNet moved object recognition past training for a finite number of categories and instead extended its intelligence to become more of an object recognition program of general intelligence (83-84).

5. What is the Mechanical Turk (Amazon's Mechanical Turk), and what role did it play in building ImageNet?

Amazon's Mechanical Turk is a service where workers can be hired to perform tasks that are simple for humans, but difficult for machines. Fei-Fei Li hired workers using Mechanical Turk to comb through the large dataset of images and label them with the proper noun for ImageNet's database. Mechanical Turk allowed ImageNet's dataset to be finished within a couple of years, as opposed to almost a century (84-85).

6. Describe the nature and operation of the ImageNet competition.

The 2010 ImageNet competition featured 1000 categories with 1.2 million images for the thirty-five competing programs to train on (85). Random categories were chosen using WordNet. Images featured one or more objects, and programs were allowed five guesses per image. The most successful program in 2010 was a "support vector machine" – it guessed 72% of images correctly (86).

7. What was the most notable thing about the 2012 ImageNet competition?

The winning program achieved 85% accuracy using a convolutional neural network, instead of a support vector machine, the popular object recognition technique of the time (86).

8. What was the most notable thing about the 2015 ImageNet competition?

One of the competitors, Baidu, cheated using a technique called "data snooping." Essentially, Baidu repeatedly used the accuracy results of the test server to improve the results of the program on the test data. While the accuracy got better each time, the results failed to show that the program could generalize, as the program merely adapted to the test data presented. While the competitors were limited to two test server submissions per week, Baidu got around this by setting up over thirty accounts to use the test server. Baidu had a short lead with 95.42% accuracy in the competition because of data snooping (88-89).

9. Describe some commercial applications of convolutional neural networks.

Image search engines improved their image recommendation software. Google Street View used convolutional neural networks to blur addresses and faces. ConvNets have also been used to automatically tag Google and Facebook images as well as filter inappropriate images on Twitter and other social networks. Self-driving cars can use ConvNet to sense pedestrians. Reading lips and body-language likewise can be done using ConvNets. Lastly, ConvNets can be used for medical diagnoses, such as breast and skin cancer (91).

10. Have ConvNets surpassed humans at object recognition?

Claims that ConvNets have surpassed humans at object recognition rely on fault-ridden premises. For one thing, all that is needed for object recognition to be a success is for the correct answer to appear in the machine's top five guesses (92). Likewise, the premise that humans error %5 percent of the time is based on an experiment with an incredibly small sample size – one human – and thus cannot be generalized (93). Additionally, ConvNets only operate accurately under specific conditions: "The kinds of errors made by ConvNets are different: while they also get confused by images containing multiple objects, unlike humans they tend to miss objects that are small in the image, objects that have been distorted by color or contrast filtres… and 'abstract representations' of objects, such as a painting or statue of a dog…" (93). Lastly, ConvNets have difficulty with localization of object recognition (94).

11. What is the relationship between "object recognition" and "visual intelligence?"

Object recognition consists of merely recognizing objects in images, whereas visual intelligence encapsulates the ability of a machine to recognize and contextualize objects based on their environment. Mitchell writes the following about visual intelligence: "If the 'objects' in question are living beings, the machines will need to know something about their actions, goals, emotions…and all the other aspects that figure into telling the story of a visual scene," (95). As it stands, artificial intelligence has a long way to go before crossing from object recognition to visual intelligence.