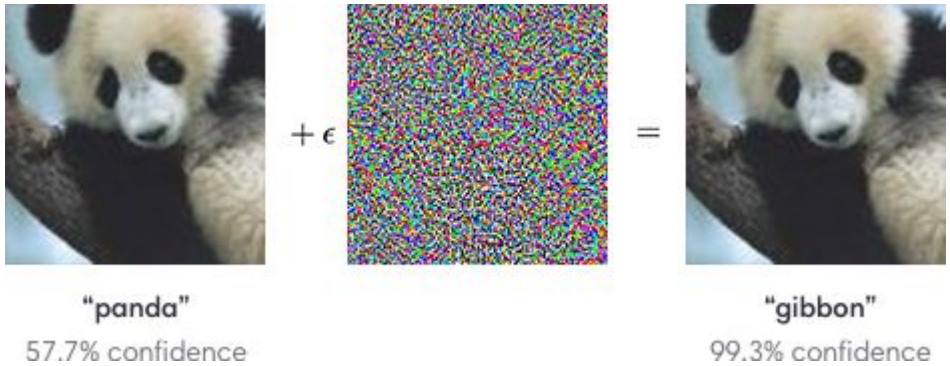
From Neural Networks to Reasoning Machines

Inherent Limitations of Deep Neural Networks

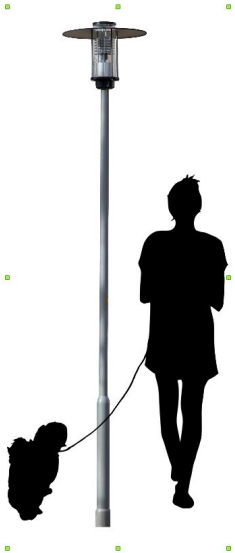
State of Deep Neural Networks (DNNs)

- The Good
  - Best solution available for image, speech and language recognition.
  - Training using DNN models is automated and it is much easier than developing algorithms
- The Bad
  - Fragile as demonstrated by Panda/Gibbon example.
  - Requires a lot of training data. A child picks up the concept of an elephant with one picture.
- The Ugly
  - DNNs Perform very badly in novel situations that were not explicitly trained for.



Perception and Reasoning, Comparison of Humans and Dogs

- Dogs
  - Dogs are nature's statisticians.
  - Unlike some robots dogs do not run into solid objects, even if they had never encountered the objects and the environment before.
- Dogs cannot project that the leash will get caught in a tree/lamp post if the post is between them and the human.
  - No amount of training will fix this because their reasoning machinery is not equipped with the necessary tools. It is also unlikely to achieve this machinery with DNNs.



- Humans
  - Not as good of statisticians as dogs
  - They have the ability to project mental models into future.
- By projecting the trajectory of the leash, they can foresee that the leash will get caught in tree/lamp post.
- Projections are made by the automatic brain system, but, since dogs cannot do that, it shows that this machinery is very specialized and cannot be achieved by a general statistical DNN.

Modeling of the environment

- Our brain recreates the environment like a holodeck.
  - When we perceive our surrounding we create a duplicate of the external objects in our consciousness. I.e., we recognize solid objects and mentally recreate them.
  - We can recreate the same model when we recall the past or create one when we are dreaming or imagining.
- .The model has many dimensions
  - Geometry
  - Solid/liquid/gas
  - Rigid/soft, temperature
  - Movement/speed

Reasoning

- Consciousness reasons by creating models and applies rules such as the laws of physics and simulates the behavior of perceived objects. Examples:
  - Two solid objects cannot occupy the same space. A dog knows that it should not collide with solid objects. As long as it correctly perceives solid objects in its environment it will not run into them. This manner of perception is very similar to ML classification plus attribute like the dimensions, or, coordinates of the solid objects. This allows dogs not requiring to be trained for different environments.
  - Human drivers project the position of their cars and moving solid objects in the environment and they can detect if these solid objects are going to collide. Once a potential collision is detected they try to solve avoiding collisions by using different alternatives in their skillset such as changing the speed of their car or swerving and changing the direction of their movement.
  - The fact that humans and dogs have different reasoning capabilities shows that there are machinery involved in their reasoning in addition to simple neural networks at work.

Using Modeled objects and laws of physics to Process Language

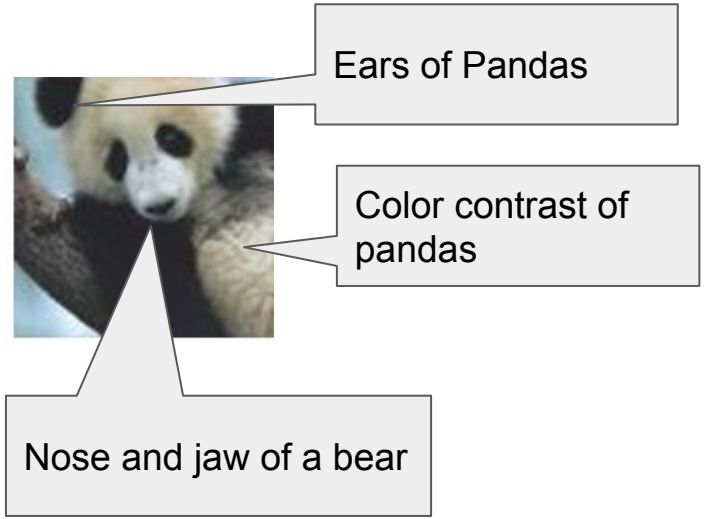
Consider the two sentences “The pen is in the box” and “the box is in the pen”

Humans can resolve this using the models for different meanings of pen. As an exercise, try searching for “pen in the box” and you will see no search engine can solve for its correct meaning.

- Writing pen is a solid object
- Box has empty space
- Pen can get into the box

- Box is inside a space
- Animal pen has space
- Box can get into an animal pen

Using Models to Classify Images



Humans construct models of objects by extracting their salient parts.

Next they break the objects they perceive into parts and match against the models they have constructed.

Their mental model of animal species resembles a tree classifier.

This hybrid approach makes human classification to need less sample data, few elephants and less prone to noise (panda/gibbon) confusion.

Summary

- Reasoning can be achieved by applying rules such as the laws of physics involving solid objects. Humans do it all the time.
- This can be applied to avoiding collisions, understanding language and classification as it was demonstrated in earlier examples.
- The first step is the current approach to machine learning, i.e., classification or recognition. Recognition is richer than simple classification, e.g., the space that the body of a solid object occupies.
- Next, apply rules such as laws of physics to the recognized objects and simulate their dynamics.