Evaluating Deep Learning Training Mechanisms for Energy Efficiency

By: Yorick Sanders, Isaac Griswold-Steiner & Dr. Abdul Serwadda

AI, Deep Learning, & Neural Networks

- **Artificial Intelligence (AI)** - a broad category describing computer systems that express human intelligence.
- **Neural Networks** - a type of AI based on neurons meant to extract features from data based on a predetermined set of outcomes.
- **Deep Learning** - A methodology of building & training networks where networks are made to extract *increasingly complex* features from data.

Neural networks are the most prominent form of AI today.

Growth of Neural Nets

- The conception of neural networks allowed computers to solve problems that only people could solve before.
- Graphics processing units and tensor processing units provided the ability to develop neural nets in minimal time, spurring the development of massive networks and allowing for powerful networks to be embedded or run on a variety of machines.

Neural Network Makeup

- Structurally, neural networks consist of **layers**.

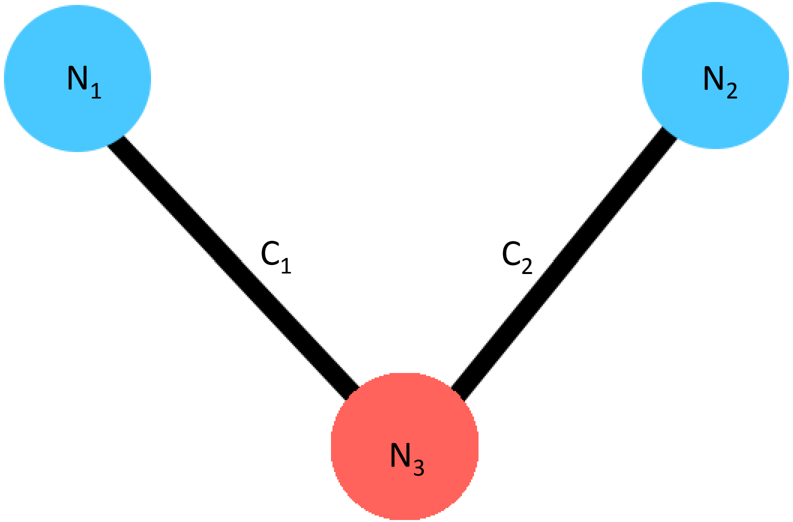
➢ Each layer has **neurons**.
  - Each neuron receives input from every neuron in the layer before it.
  - Each neuron gives output to every neuron in the layer after it.

➢ Connections have “strengths” that dictate how much of an effect data passing through the connection has on subsequent neurons.
  - Neural networks are “taught” by modifying the strengths of their connections to mold the network into producing more desired outputs.

Network Structure

A simple neural network:

- Networks can be **trained** and **tested**.
- Testing is giving a network some input (like an image) and running the network, so that it produces an output (e.g., “Dog”).
- Training is actually modifying a network so that it gives more desired outputs when tested.



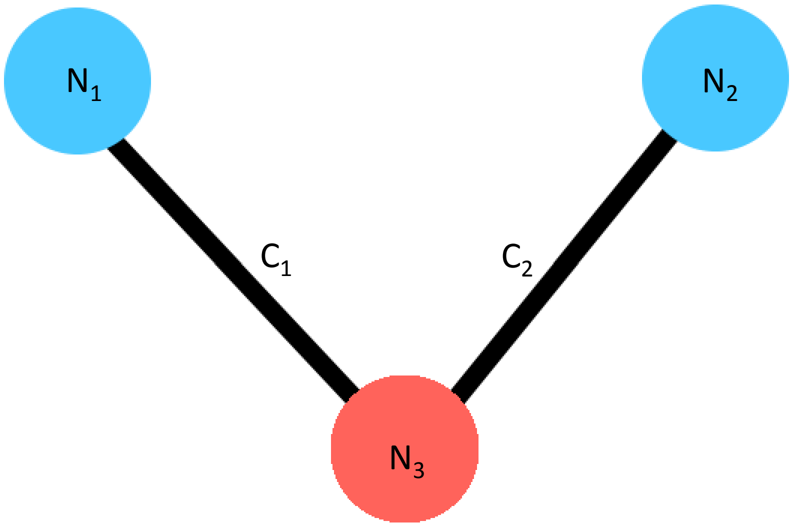
- During testing, each blue neuron sends a number down the connections, to the red neuron.

- The connections have numeric values that affect the numbers being sent through them in some way.

Network Structure

A simple neural network:

- During training, the network is tested and then adjusted to produce a more desirable output.
- **The process of training a network involves testing it.** This is one of the reasons that training typically takes more power.



Calculating the Output

- Sum the multiples of the weights times the inputs.
- Add the bias
- Apply the activation function
- Do this for every neuron (can be millions or more)

Testing

$$output = f(b + \sum_{i}^{n} x_i w_i)$$

Training

$$\frac{\partial L}{\partial u_{ij}} = \frac{\partial L}{\partial y_j} \frac{\partial y_j}{\partial a_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial u_{ij}} = (y_j - t_j) y_j (1 - y_j) h_i = \delta_j h_i$$

Energy Consumption

- Some types of networks can require lots of power.
- High energy requirements for Neural Network operations.
    - Training a single AI can take about 284 tons of CO_2 equivalent – about as much as the lifetime of 5 cars![1]
- Large companies train tens of thousands of models every year.

Scenario	CO2 (lbs)
Human life (avg) 1yr	11,023
American life (avg) 1yr	36,156
Car, avg incl. fuel, 1 lifetime	126,000
NLP pipeline (parsing, SRL) * w/ tuning & experiments	39 78,468
Transformer (big) * w/ neural architecture search	192 626,155

Purpose

- This research aims to determine the impact of different means of training on energy consumption.
- Approaches such as transfer learning and network pruning may reduce energy consumption, but we must also evaluate the impact on task performance.
- Additionally, the energy consumption of different common training operations (like normalization and augmentation) is measured.

Transfer Learning

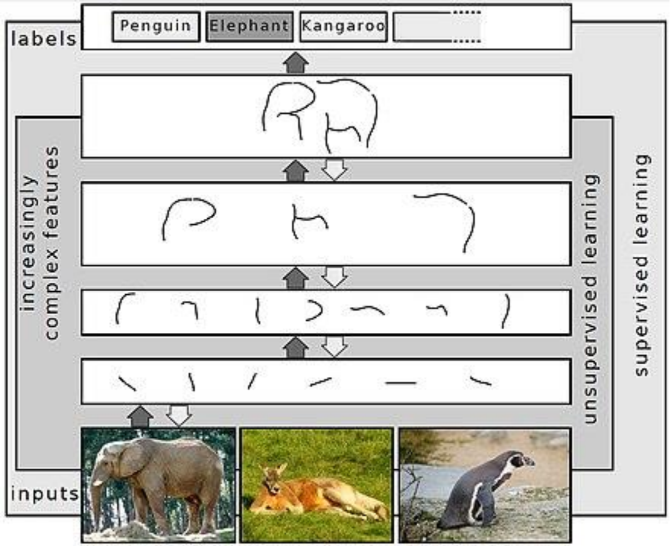
Consider the following scenario:

- You're given two networks, **Network A**, that is already trained to tell you each body part, and its location, of any animal in an image you give it, and **Network B** that isn't trained to do anything yet.
- You need to develop a network to recognize whether an image contains a dog.
- Which network would you start with to get a dog-recognizer with minimal effort?

Answer: Pick **Network A**, add another few layers, and do a little more training so its final output is the presence of a dog.

Why Transfer Learning?

Network A might not have been trained *just* to recognize dog parts in an image, but it’s still a better option than training a network from scratch because most of the work is already done; the input to the part you need to train is now animal parts with location information, rather than raw images, which is information that is significantly closer to the goal.

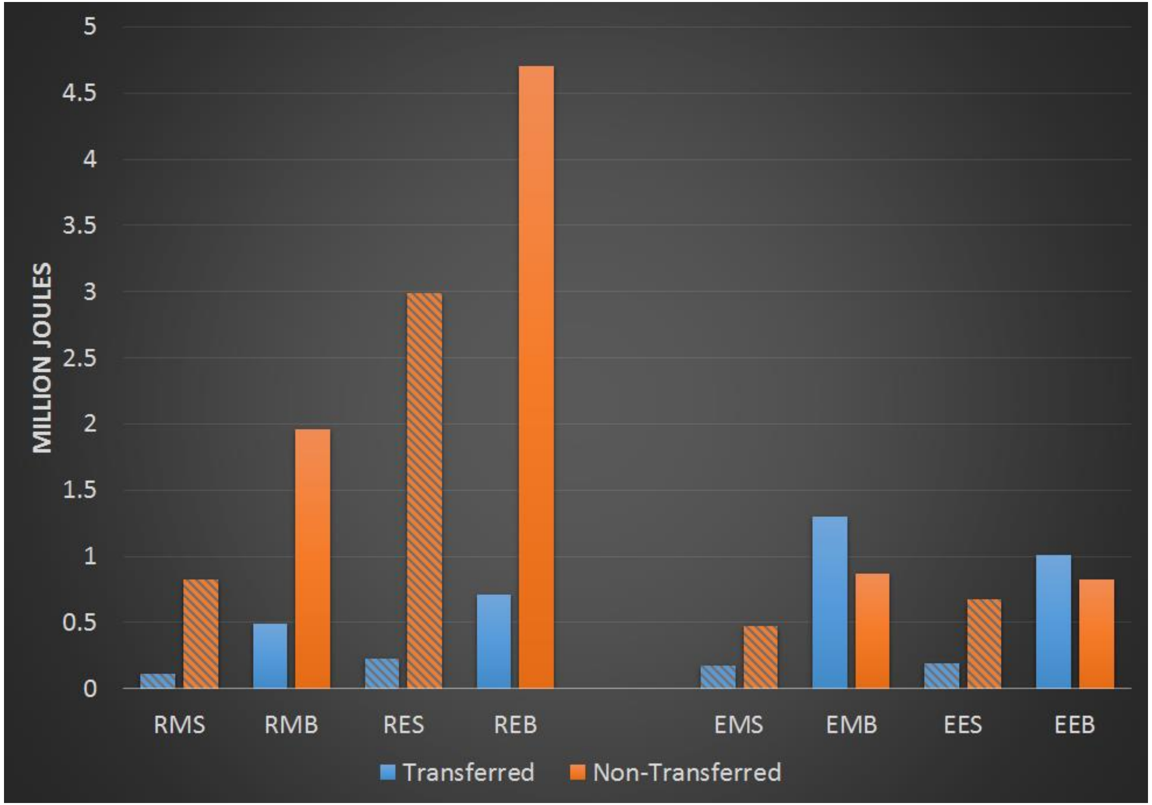


Taken from: https://www.wikiwand.com/en/Deep_learning

Methodology

- We took phone data of users performing actions with phones in their pockets and trained networks to recognize which person each datum belonged to.
- Data was taken from the Extrasensory and Motion Sense datasets.
- We compared results between a large network and a small network.
- Energy usage between both training and testing of transferred and non-transferred networks was measured heretofore.

Energy Consumption Results



1st letter:

- R = Training
- E = Testing

2nd letter:

- M = Motion Sense
- E = Extra Sensory

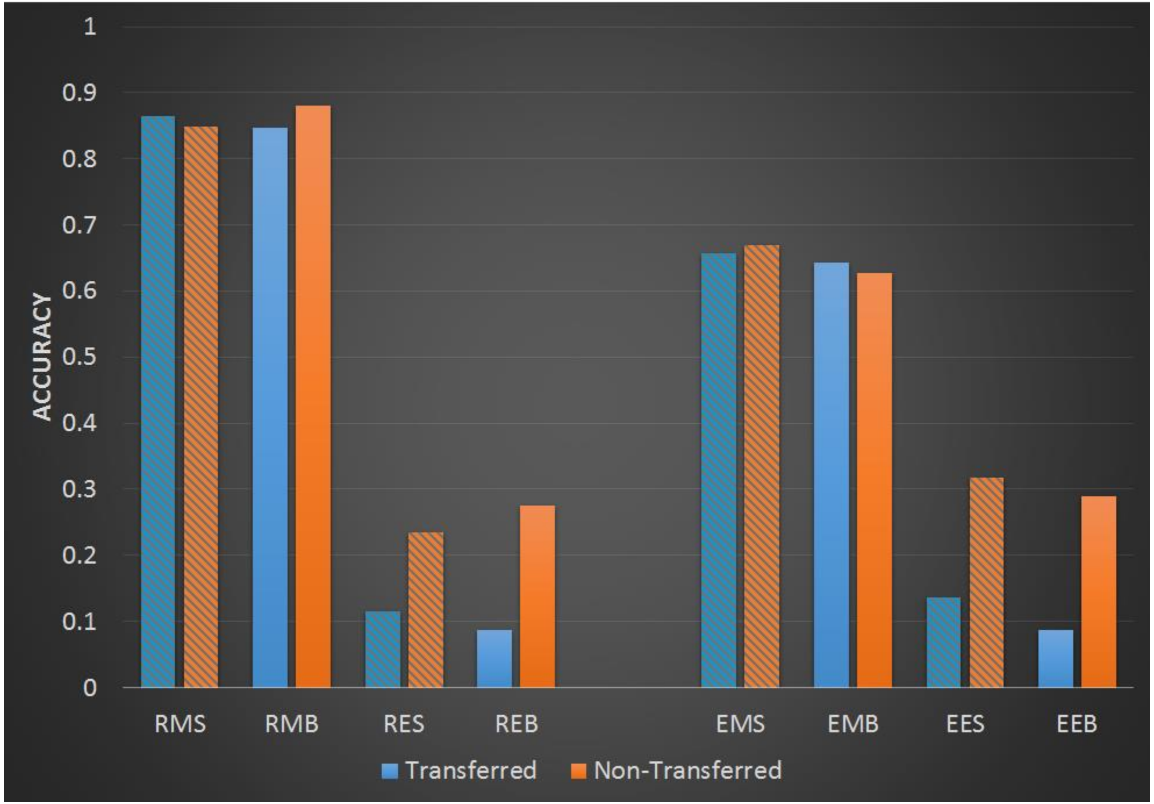
3rd letter:

- B = Big Network
- S = Small Network

Energy Consumption Analysis

- During training, the transferred network consumed between 7% and 26% of the energy that the non-transferred network used.
- During testing, energy consumption was between 20% and 50% higher for the big transferred network than the big non-transferred network. For the small network during testing, the transferred network was more efficient.

Accuracy Results



1st letter:

- R = Training
- E = Testing

2nd letter:

- M = Motion Sense
- E = Extra Sensory

3rd letter:

- B = Big Network
- S = Small Network

Accuracy Analysis

- Accuracy performance for the transferred network was on par with the non-transferred network for the Motion Sense data set (differing by up to 5%). The transferred network however performed between about 1/3rd to half as well as the non-transferred network for the Extra Sensory dataset.

Conclusion

Using transfer learning for the Motion Sense dataset during training allows for a significant reduction in power consumption while maintaining a reasonable level of accuracy.

This approach can be viable if the network is of a small scale (where the transferred network used less energy) or when testing isn't performed very frequently and the network is large.