Deep learning and hierarchical temporal memory
The search for machine intelligence

A Thesis submitted

in Partial Fulfilment of the Requirements

for the Degree of

Bachelor of Technology

by

Vinay Varma

(141200043)

Department of Electronics and Communication Engineering

NATIONAL INSTITUTE OF TECHNOLOGY, DELHI

2017- 18

APPROVAL SHEET

This project work entitled

Deep learning and hierarchical temporal memory :The search for machine intelligence

by Vinay Varma is approved for the degree of

Bachelor of Technology

Examiners:

_____

_____

Supervisor(s):

_____

_____

Chairman:

_____

Date:

Place:

DECLARATION

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/ source in my submissions. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal actions from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Vinay Varma

141200043

Department of Electronics and Communication Engineering

Supervisor:

Dr. Baljit Kaur

Assistant Professor

Department of Electronics and Communication Engineering

Date :

Abstract

In the summer of 1956 a group of researchers gathered on the campus of Dartmouth College.It is the first workshop on Artificial Intelligence.It was then declared that a machine as intelligent as a human being would exist in no more than a generation.More than 60 years later we do not yet even have a general consensus on the definition of intelligence.We have soon realized the difficulty of the problem.Here i provide a brief survey of possible paths towards A.G.I.I would argue how the architecture of the brain is fundamentally different from the modern state of art machine learning approaches.To discuss the short-comings of the present models i have chosen to take the task of image recognition.Here i have trained a deep convolutional network to classify between cats and dogs.As i wrote the problem of A.G.I is extremely difficult, if not intractable.Here i would evoke the ideas from mathematics,computer science,physics,neuroscience and evolutionary biology in my search for machine intelligence.

Synopsis

Names of the Student	:	**Vinay Varma**
Roll Number	:	**141200043**
Degree for which submitted	:	**Bachelor of Technology**
Department	:	**Electronics and Communication Engineering**
Thesis Title	:	**Deep learning and hierarchical temporal memory The search for Machine Intelligence**
Thesis Supervisor	:	**Dr. Baljit Kaur**
Month and year of submission	:	**April, 2018**

The thesis has been organized in four chapters, the summary of which is given below:

Chapter 1 defines the learning problem.

Chapter 2 discusses the actual algorithm.It discusses Resnet32 architecture that we have used for object recognition.

In Chapter 3,We put forth the shortcomings of the present models.The major thesis of the project is discussed here.

In Chapter 4 ,We conclude our work,with some thougghts on the future of Artificial Intelligence.

The relevant references are appended at the end.

Acknowledgement

I would like to express my deep gratitude to **Dr.Baljit Kaur**, Assistant Professor, ECE Department, for her expert guidance and enthusiastic encouragement throughout, from the commencement of the project to its completion.

I am profoundly grateful to all the faculty and non-faculty staff of the department of Electronics and Communications who helped us during the course of this project. Finally we thank our parents and friends for the much needed moral support and to whom we owe everything.

(Vinay Varma)

Contents

Approval sheet **i**

Declaration **ii**

Abstract **iii**

Synopsis **iv**

Acknowledgement **v**

List of Figures **viii**

Acronyms and Abbreviations **ix**

1 Modeling human intelligence with probabilistic program induction **1**

1.1 Objective 1

1.1.1 The Learning problem 1

1.1.2 What they have shown ? 2

1.1.3 What is BPL ? 3

1.1.4 The details and the results 4

1.1.5 The not so cool parts of the paper 7

1.1.6 *Extra* 9

2 The Experiment **11**

2.1 The Model 11

2.1.1 The design and Evolution of Convolutional Neural Networks 11

2.1.2 The Problem 12

2.1.3 Implementation 14

2.1.4 Results 16
2.1.5 Future Work 16

3 The Hypothesis **17**

3.1 Where did we go wrong ? 17
3.1.1 On Brain 17
3.1.2 Science and Simple Explanation to Complex Phenomenon 18
3.1.3 Rocks and Brains: Why We Might Not Be Special ? 19
3.1.4 Why Our Brains Are Not Optimal designs ? 25
3.1.5 The Evolution Of Brain 28

4 The Conclusion **30**

4.1 So,what's wrong with the current approach ? 30
4.1.1 The Future Ahead 31
4.1.2 Law Of Accelerating Returns 31

Bibliography **35**

List of Figures

1.1 2
1.2 Symbolic era 9

2.1 Resnet34[1] 12
2.2 Residual Learning:a building block 12
2.3 Some images in the dataset 13
2.4 Cyclical learning rates[2] 15

4.1 Brain Scanning Technology[7] 32
4.2 Exponential trend in the increase of order and complexity.[7] 33
4.3 Compute power vs cost[7] 34

Acronyms and Abbreviations

A.G.I Artificial General Intelligence

Chapter 1

Modeling human intelligence with probabilistic program induction

1.1 Objective

Here i would try to provide a broader perspective on the authors idea.Tenenbaum has been working on the core idea of this paper for quite some time - compositionality, causality, and learning to learn.Though i tried to stick with the authors present work,i have to draw considerably from his other works too.(Building Machines That Learn and Think Like People - 2016).Wherever i couldn't completely digress to my heart's content(space constraints) i have cited the original resource from which i learned as i think it's important to deeply appreciate certain ideas to see where a new idea fits into the fabric.If the line of ideas presented here seems out of place they may either be so or a brief pondering would fit them onto the edifice.

1.1.1 The Learning problem

If we look at the theoretical basis of machine learning we see that there is this trade off between model complexity and generalizability.We need more data for a more complex model to generalize well - (captured by vc dimension or much simpler bias-variance trade off).But humans seems to do one shot learning.Authors objective was to achieve human

level performance in a task not through traditional statistical approach(refer *extras*(1)) but from the understanding of how humans do it ?

1.1.2 What they have shown ?

Authors have shown that through BPL(bayesian program learning) the engine can learn to perform one shot learning on par with humans.Following picture outlines their achievement..

By looking at the segway we can easily identify it's compositional parts - we can iden-

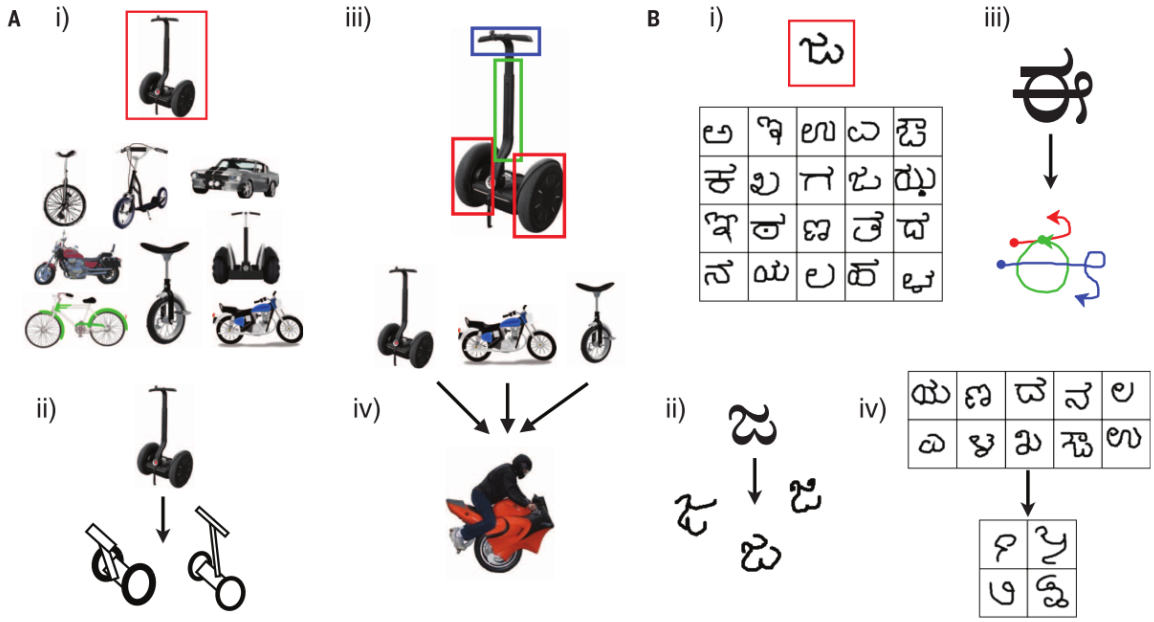


Fig. 1. People can learn rich concepts from limited data. (**A** and **B**) A single example of a new concept (red boxes) can be enough information to support the (i) classification of new examples, (ii) generation of new examples, (iii) parsing an object into parts and relations (parts segmented by color), and (iv) generation of new concepts from related concepts. [Image credit for (A), iv, bottom: With permission from Glenn Roberts and Motorcycle Mojo Magazine]

Figure 1.1

tify the wheels,suspension etc,draw it,generate new models by combinig with previous similar models.They have taken various hand written letters from various languages and tested their engine on 4 different tasks that involves learning a concept from single example,generating new examples,parsing whole into parts and relations and generating new concept from old ones.The detail of the engine are omitted in the paper though i would provide some of the details in the next section.But it has to (a) build causal models of the

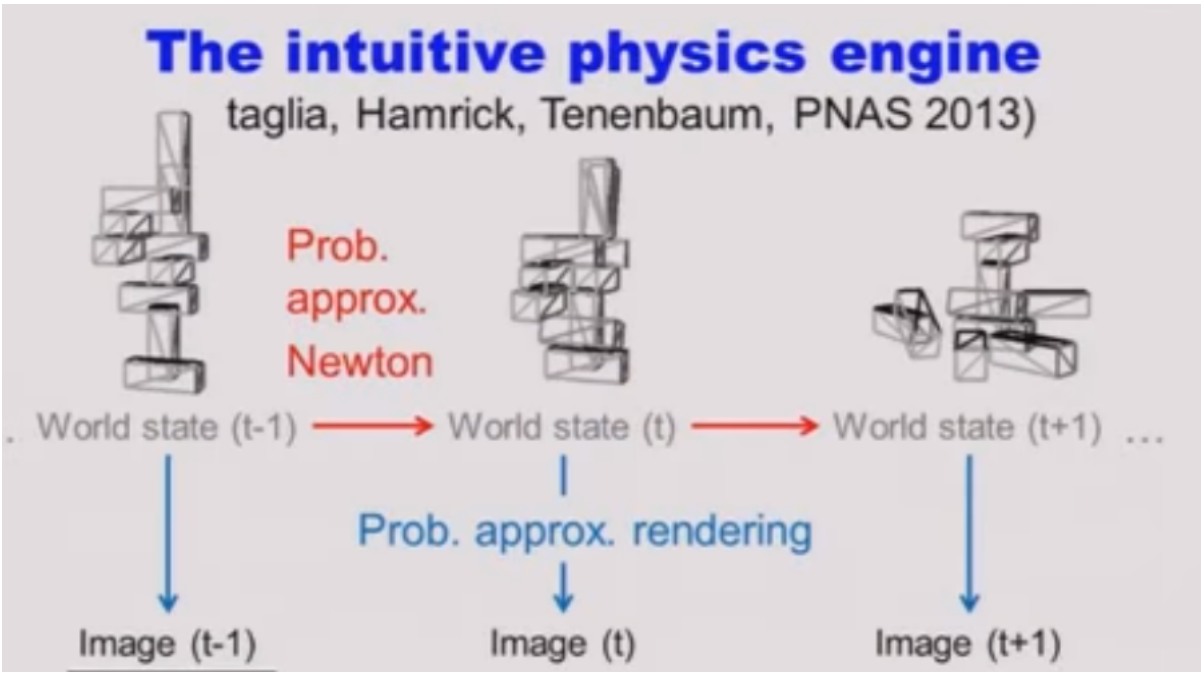
world that support explanation and understanding, rather than merely solving pattern recognition problems; (b) ground learning in intuitive theories of physics and psychology, to support and enrich the knowledge that is learned; and (c) harness compositionality and learning-to-learn to rapidly acquire and generalize knowledge to new tasks and situations.

They have incorporated inductive learning quite successfully through BPL- show a new letter - some body has drawn it - It must be sequence of strokes - strokes match most closely to 'A' - so it must be A.

1.1.3 What is BPL ?

Bayesian Program Learning (BPL) represents concepts as simple stochastic programs that is, structured procedures that generate new examples of a concept when executed. These programs allow the model to express causal knowledge about how the raw data are formed, and the probabilistic semantics allow the model to handle noise and perform creative tasks. Structure sharing across concepts is accomplished by the compositional reuse of stochastic primitives that can combine in new ways to create new concepts.
Rather than cluttering the writeup with the details of the experiment let's try to intuitively understand what's happening under the hood.We have probabilistic graphical models(Daphne kollers book and coursera course are amazing resource) that helps us to build general purpose causal models.Probabilistic programs are just more expressive versions of pgm's.HMM'S are great at this task but they miss much of the structure just like how flow charts miss much of the structure in a traditional computer program.A word isn't just a finite dimensional vector nor physics can be a transition matrix.Instead in BPL authors have used seperate programs to induce intuitive physics and psychology.They are built on two core insights:
1.Humans come with intuitive prior's - like an understanding of physical objects,intentional agents,their properties etc..
2.Learning is theory building(child as a scientist not data analyst.) BPL has been built by many scholars under tanenbaum for instance the physicse engine(which closely parallels

with game engines).I would also suggest to read Deep Convolutional Inverse Graphics Network - tanenbaum to understand further details.Similarly an intuitive psychology engine is also built in .We know through evolution enduring properties of the world are bulit in as priors,we don't come with blank slate.Then it is sensible to build corresponding priors into the algorithm.Some technical details and the results are covered in next section.



1.1.4 The details and the results

The following details are not directly mentioned but are very crucial to understand:
Authors have assumed the primitives(below figure) before performing the experiment.Notable point is that their accuracy is not so crucial to the performance of the algorithm as the sequence of strokes is learned by the algorithm which is more crucial for the accuracy.In one setting of the experiment,the algorithm is provided with a new example that is not in the training data,the compositional parts are inferred and then based on a similarity metric it is classifed appropriately.So,as assumed we are give basic elements like a turn,hook,wave etc called primitives.When these primitives are connected sequentially parts are formed - those that can be drawn with a single stroke.Here are the most crucial parts of the algorithm:

Search space of the algorithm is restricted in the following two ways: 1.The fact that we are drawing single stroke line from primitives constrains their location. 2.Now consider two stroke letters- like the combination of epsilon and the hook.We have three line ends there.Now by varying relative position and adding motor noise we generate different samples.It's important to notice how we are limiting search space here.

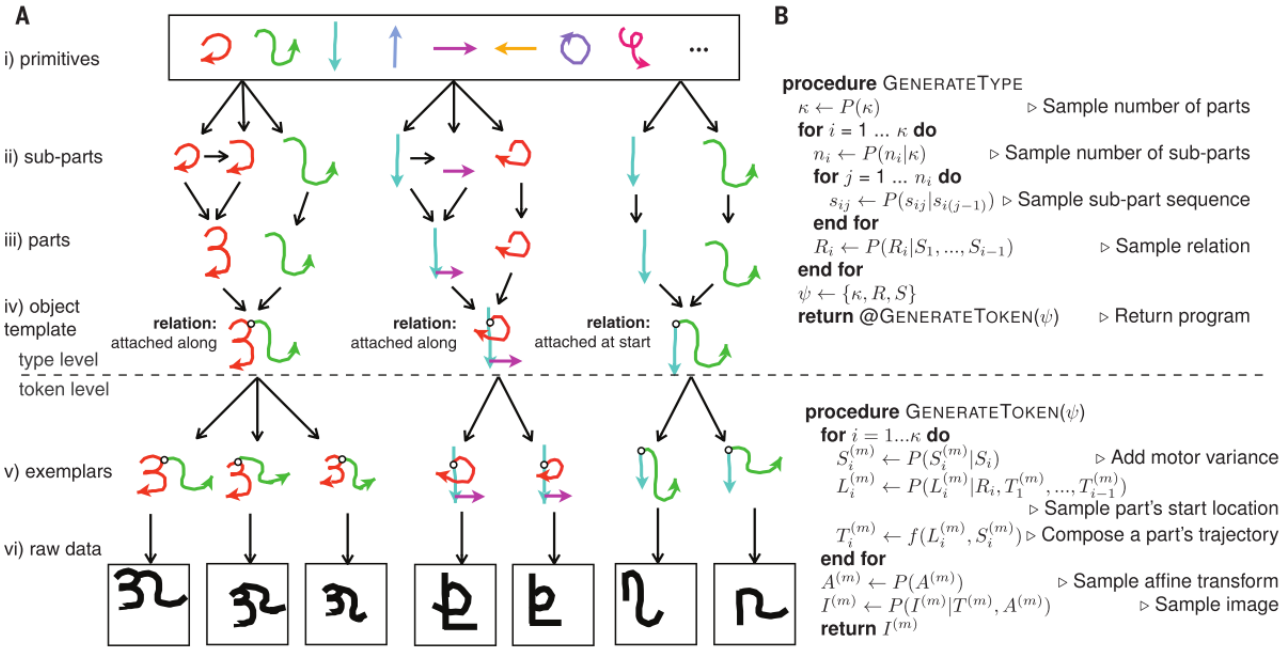


Fig. 3. A generative model of handwritten characters. (**A**) New types are generated by choosing primitive actions (color coded) from a library (i), combining these subparts (ii) to make parts (iii), and combining parts with relations to define simple programs (iv). New tokens are generated by running these programs (v), which are then rendered as raw data (vi). (**B**) Pseudocode for generating new types ψ and new token images $I^{(m)}$ for $m = 1, ..., M$. The function $f(\cdot, \cdot)$ transforms a subpart sequence and start location into a trajectory.

Note that BPL is a generative model so there can be multiple programs with different parses but still that isn't a problem as illustrated..

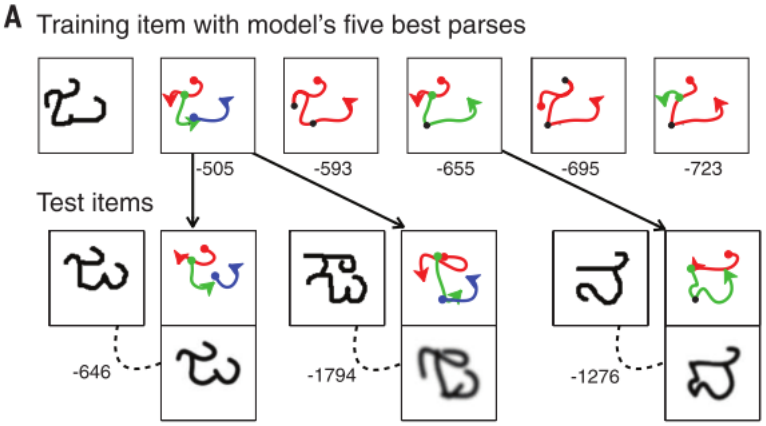


Fig. 4. Inferring motor programs from images. Parts are distinguished by color, with a colored dot indicating the beginning of a stroke and an arrowhead indicating the end. (**A**) The top row shows the five best programs discovered for an image along with their log-probability scores (Eq. 1). Subpart breaks are shown as black dots. For classification, each program was refit to three new test images (left in image triplets), and the best-fitting parse (top right) is shown with its image reconstruction (bottom right) and classification score (log posterior predictive probability). The correctly matching test item receives a much higher classification score and is also more cleanly reconstructed by the best programs induced from the training item. (**B**) Nine human drawings of three characters (left) are shown with their ground truth parses (middle) and best model parses (right).

RESULTS:Most notable part of the results is that BPL with leisons have dramatically increased error rate.It shows that by adding additional constraints by providing the knowledge of how the characters are made search space is reduced thus decreased error rate.

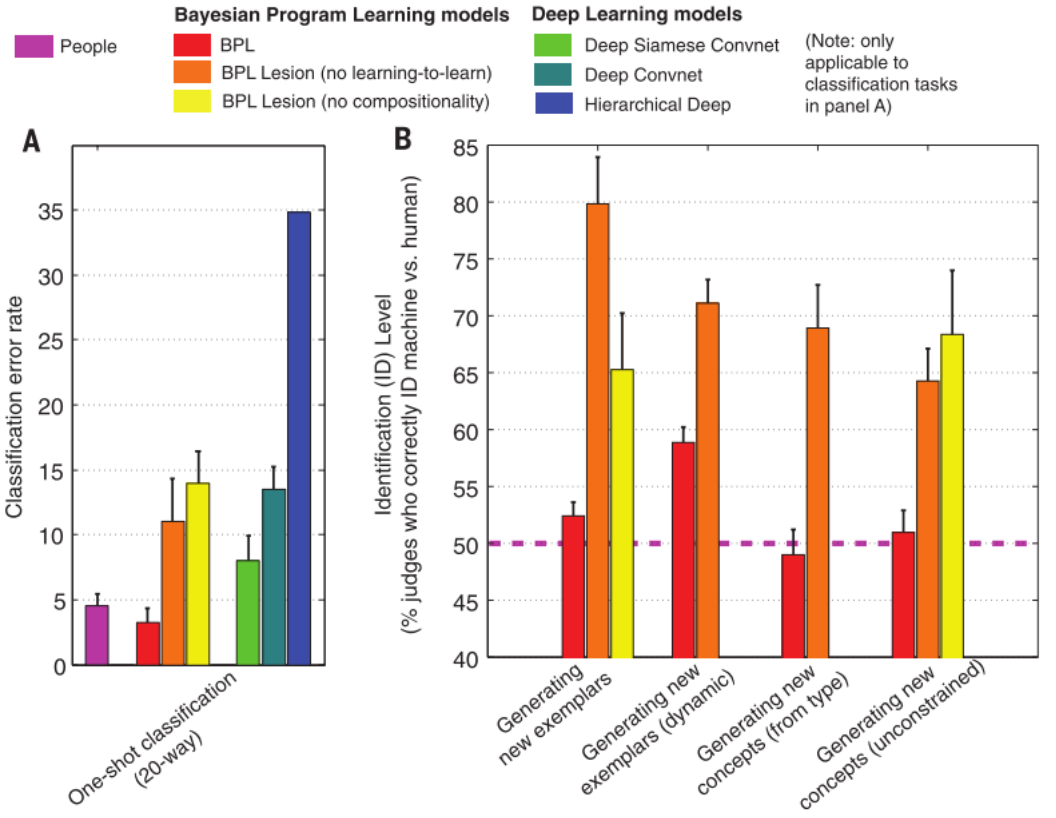


Fig. 6. Human and machine performance was compared on (A) one-shot classification and (B) four generative tasks. The creative outputs for humans and models were compared by the percent of human judges to correctly identify the machine. Ideal performance is 50%, where the machine is perfectly confusable with humans in these two-alternative forced choice tasks (pink dotted line). Bars show the mean ± SEM [*N* = 10 alphabets in (A)]. The no learning-to-learn lesion is applied at different levels (bars left to right): (A) token; (B) token, stroke order, type, and type.

1.1.5 The not so cool parts of the paper

1.The notion of providing inductive bias is sensible but authors have cleverly chosen the right problem setup to easily come up with right primitives(they have used deep learning).Consider language learning or visual scene understanding - what are the right primitives ? 2.For instance the authors haven't included line width into their basic primitive structure.Same sequence of strokes but different line widths could result in different estimation of the probabilities. 3.According to authors own note, BPL, sees less structure in visual concepts than people do. It lacks explicit knowledge of parallel lines, symmetry, optional elements such as cross bars in 7s, and connections between the ends of strokes and other strokes. 4.Now,here's the biggest drawback.It seems that the smartest people in the field are asking the wrong question.Firstly,We have to understand

that humans are born with strong priors that are geared towards our survival in this universe.Under evolutionary time scale what has been constant - laws of physics and basic human nature has provided us with strong prior.We are not as fast learners as the A.I community led us to believe.If we are to put under slightly different laws of physics then i would bet we are far from one shot learners.A recent paper clearly drives this point- INVESTIGATING HUMAN PRIORS FOR PLAYING VIDEO GAMES(uc berkeley).

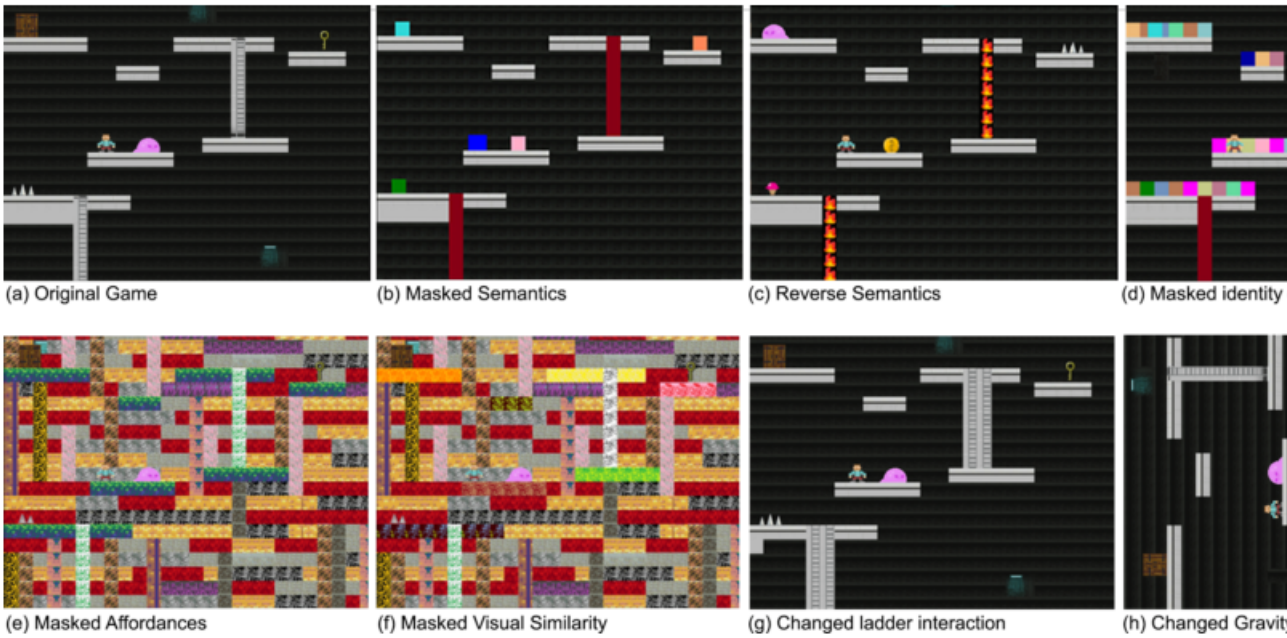


Figure 2: **Various game manipulations**. (a) Original version of the game. (b) Game v objects to ablate semantics prior. (c) Game with reversed associations as an alternate w semantics prior. (d) Game with masked objects and distractor objects to ablate the conce (e) Game with background textures to ablate affordance prior. (f) Game with backgrou and different colors for all platforms to ablate similarity prior. (g) Game with modifi hinder participant's prior about ladder interactions. (h) Rotated game to change partic about gravity.

A few minor changes like gravity direction has multiplied the learning time by few decades.Being equipped with strong prior knowledge can sometimes lead to constrained exploration that might not be optimal in all environments (Lucas et al., 2014; Bonawitz et al., 2011).Investigating human priors for every task and incorporating that into A.I is no dumber than rule based systems.

1.1.6 *Extra*

The following text is not part of the critique but this serves to provide additional context to the arguments in the text.

(1)I think we can break the history of A.I into three eras...

1.Prehistory - 1980 Symbolic A.I(Minsky being the major proponent).Largely over influenced by the success of the computers and drawing analogs to the brain people believed that brains are mere symbol manipulating machines.

2.Intelligence as statistics on large scale (1980's to 2000):Given enough data any function

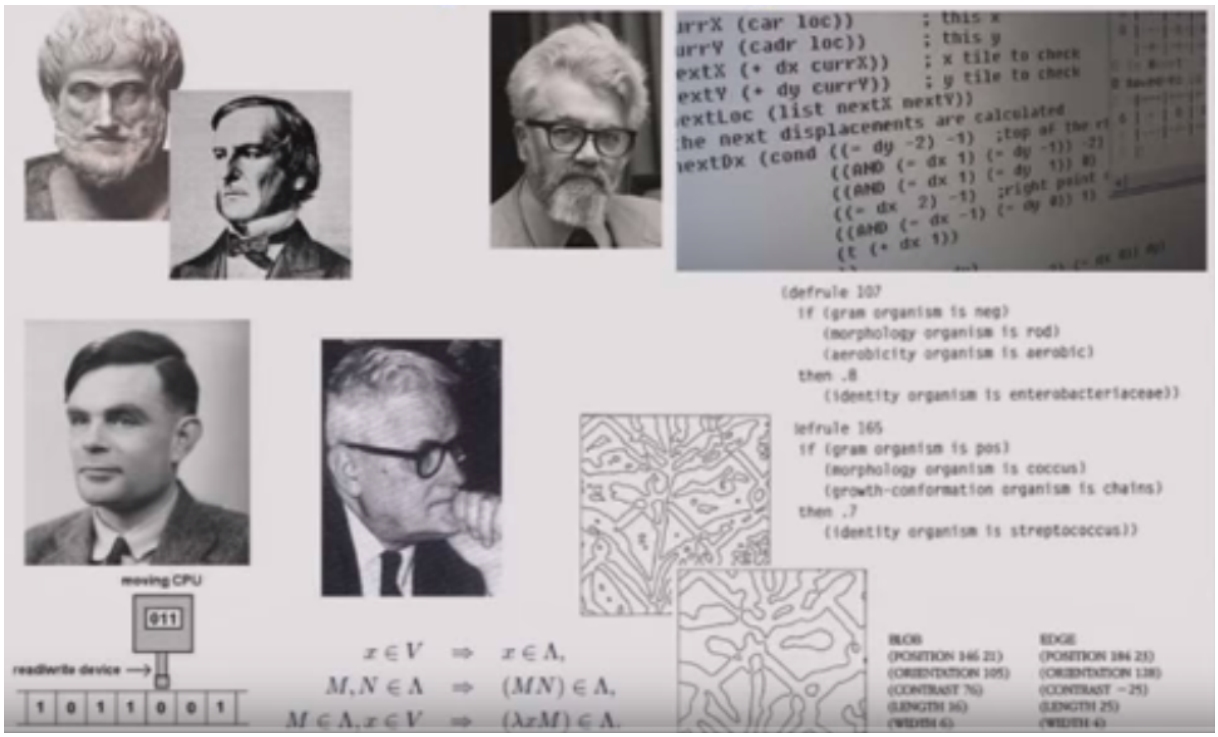


Figure 1.2: Symbolic era

can be estimated,so our brains must be doing data analysis.Much of the math of finding structure in the data(deep learning to spectral clustering methods) is worked out sort of underground in 60's but burst into the scene from 80's.

3.Present:We have made special purpose algorithms to beat human performance in several tasks.Consider state of computer vision - we have outperformed humans in object localization tasks but vision isn't just about that.When we see an image we also infer many intuitive things like the mental state of the person,possible future evolution of the

system,social relationships of the people from their expressions etc.In fact this inability is one of the biggest impediment to the development of self-driving cars - visual scene understanding.People have realized that just by combining ideas from the previous two eras is not the way to go.Researchers like tanenbaum work on the intersection of cognitive science and A.I to develop more human like machines.

Chapter 2

The Experiment

2.1 The Model

Here we will look at the model i have used to develop the image classifier.This particular architecture is first proposed by KaimingHe[1] and his research group Deep Residual Networks For Image Recognition.This is the current state of the art machine learning model for image recognition.

2.1.1 The design and Evolution of Convolutional Neural Networks

The above diagram shows the design of three popular architectures.The last one resnet-34 is the one we have used to design the classifier.This primary advantage of the above architecture is that mapping identity function becomes easier.The skip connections indicated in the figure by the curved arrows shows the addition of signals.i.e ,

In deep learning,as discussed in the previous chapter we face the problem of over-fitting vs under-fitting.If the model is too complex,it would immediately over-fit.So,traditionally models with lot's of layers are great at capturing a more complex problem like vision.So,deeper networks generally better,if provided with a method of regularization.This refers to the control of the programmer to constrain the search space of the model.For

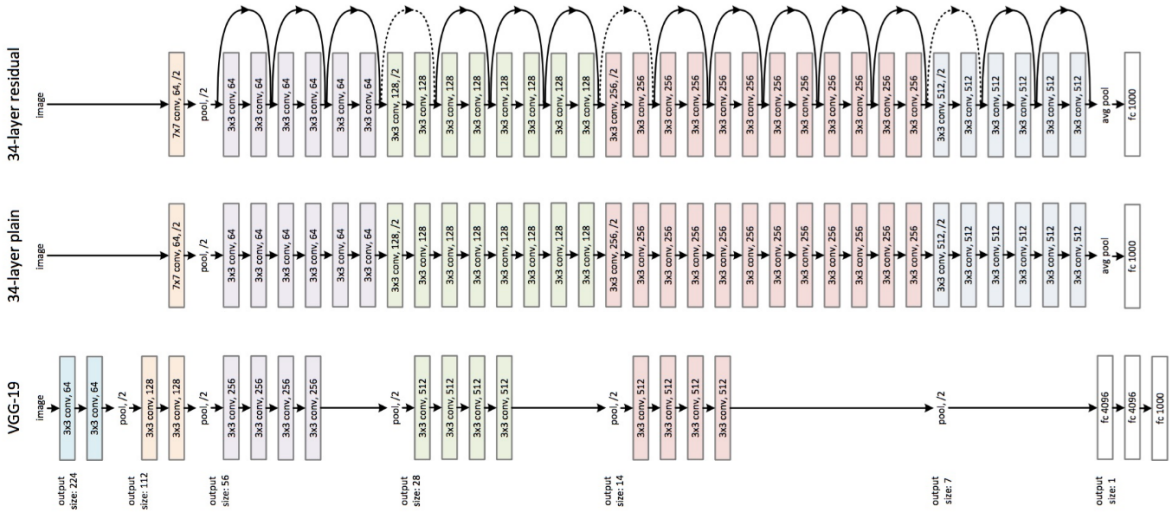


Figure 2.1: Resnet34[1]

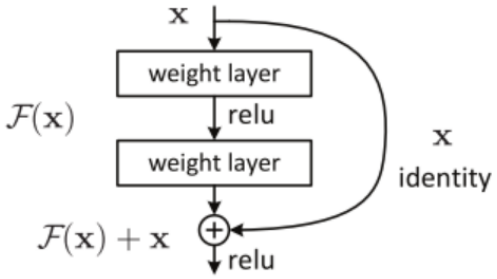


Figure 2.2: Residual Learning:a building block

instance if we constrain of model to have values of the parameters w between (0,1),that is an example of Regularization.This, when formalized the community has come up with several methods some of which i would try to elaborated in the previous chapter.

2.1.2 The Problem

Here we are given images of 120 different dog breeds.For each breed we are given 50 to 160 images.Using this information we are asked to create a dog breed classifier that could take input a dog and output it's breed.Before actually attempting the problem i have tried to solve a much simpler problem where i tried to classify between cats and dogs given images of each.I will summarize the results of it before proceeding further.

(a) A random dog (b) A random cat

Figure 2.3: Some images in the dataset

The images in the above diagram shows some of the correctly identified images.While the images below shows the ones the network has confused the most.A probability of 1 represents a dog while 0 shows a cat.

2.1.3 Implementation

The architecture defined above is implemented using various software packages,they are listed below:

1. Python : A general purpose programming language.
2. Pytorch : A package to implement and debug deep learning models.
3. Fastai : A wrapper around Pytorch.
4. Numpy :A numerical programming tool.
5. Pandas : A tool for visualizing data and handling missing values.

Here we would provide some other notable details of the implementation.

1. Choosing a learning rate: The learning rate determines how quickly or how slowly you want to update the weights (or parameters). Learning rate is one of the most difficult parameters to set, because it significantly affect model performance.In the program we have used the method $learn.lrfind()$ which finds an optimal learning rate. It uses the technique developed in the 2015 paper Cyclical Learning Rates for Training Neural Networks, where we simply keep increasing the learning rate from a very small value, until the loss stops decreasing. We can plot the learning rate across batches to see what this looks like.
2. Data Augmentation:If you try training for more epochs, you'll notice that we start to overfit, which means that our model is learning to recognize the specific images in the training set, rather than generalizaing such that we also get good results on the validation set. One way to fix this is to effectively create more data, through data augmentation. This refers to randomly changing the images in ways that shouldn't impact their interpretation, such as horizontal flipping, zooming, and rotating.We

can do this by passing $augtfms$ (augmentation transforms) to $tfmsfrommodel$, with a list of functions to apply that randomly change the image however we wish. For photos that are largely taken from the side (e.g. most photos of dogs and cats, as opposed to photos taken from the top down, such as satellite imagery) we can use the pre-defined list of functions $transformssideon$. We can also specify random zooming of images up to specified scale by adding the $maxzoomparameter$.

3. Cyclical Learning Rate:I used a technique called stochastic gradient descent with restarts (SGDR), a variant of learning rate annealing, which gradually decreases the learning rate as training progresses. This is helpful because as we get closer to the optimal weights, we want to take smaller steps.

   However, we may find ourselves in a part of the weight space that isn't very resilient - that is, small changes to the weights may result in big changes to the loss. We want to encourage our model to find parts of the weight space that are both accurate and stable. Therefore, from time to time we increase the learning rate (this is the 'restarts' in 'SGDR'), which will force the model to jump to a different part of the weight space if the current area is "spikey". Here's a picture of how that might look if we reset the learning rates 3 times (in this paper they call it a "cyclic LR schedule"):

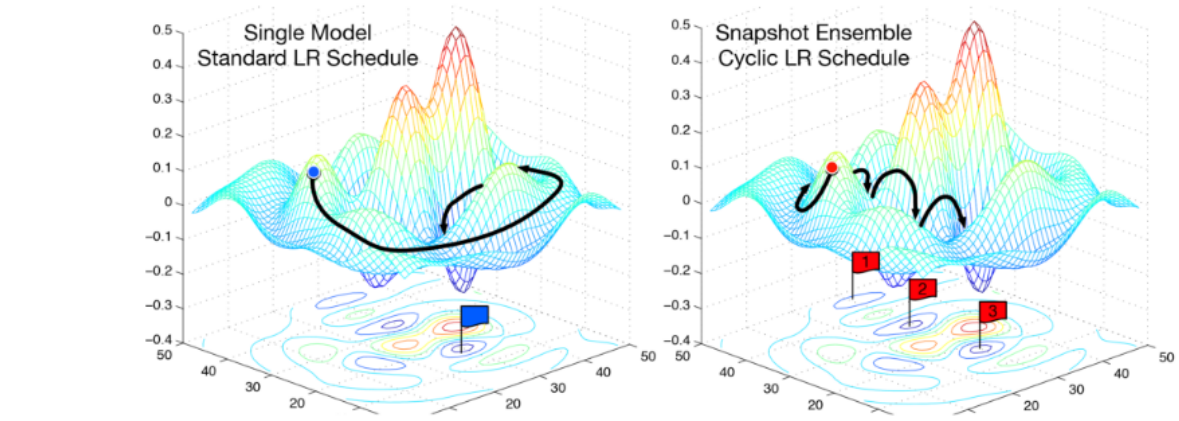


Figure 2.4: Cyclical learning rates[2]

2.1.4 Results

After running the model for several epochs we got the accuracy of 92.3.We have designed a dog classifier that can be used to identify any dog-breed.

2.1.5 Future Work

This same model can be used for general purpose training of any image classifier.Apart from this only with a few minor modifications we can make the popular face recognition applicationss like the one in iphone x.But,this is not the motive of the thesis as described earlier.In next chapter we will explain why the model which is current state of art lacks the generalization capacity of human visual system.

Chapter 3

The Hypothesis

3.1 Where did we go wrong ?

In the previous section we have created a state of the art image classifier.But,on careful examination we can see that how the algorithm is using substantially large amount of data compared to the human subject.In my project the algorithm is using about 100 examples of each class through 100's of iterations which amounts to 10000 examples before actually recognizing the correct object.Apart from that we have shown how the algorithm mistakes a cat for dog which a human subject is highly unlikely to commit.This also indicates how our learning process might be fundamentally different from the algorithm mentioned in the previous chapter.In this section we will look at the fundamental mechanism of human learing and their genetic roots.

3.1.1 On Brain

We have learned very little about our brains.May be like all complex systems we will one day figure out its working from how we form opinions to our ability to be self-aware.We agree that it is genuinely puzzling how interactions within a 3 pound matter can produce theories of cosmos,be self aware and even speculate on its origins.But so are the vast number of phenomenon in the universe–we have managed to understand and harness

them to our benefit.Then question pertains–does the brain too operate on same physical laws.If that is the case then we would argue–we are very close in making our very own brain(maybe without some of its glitches)–that would do all our inventions for us.If that needs us to extend physical laws–thats okay we would as long as those dont tells us of a master with a magic wand.

3.1.2 Science and Simple Explanation to Complex Phenomenon

Science is a breathing book.We have been changing our notions of reality.All through our history our intuition about the workings of nature have been failing us.We are not the centre of the universe.We are not a separate creation from all the other animals.We are not here by any purpose.We are a phantom of chance.

> Calvin in the comic strip Calvin and Hobbes.: The best proof that there is intelligent life in the universe is that they have not contacted us. Maybe its because they would die laughing.

Probably the most striking lesson we have learned through science is that all the complex phenomenon we see can be accounted by surprisingly simple facts.Apparent complexity is the manifestation of long chains of simplicity.This shouldnt come by as surprise for in a perfectly godless world how can something complex exist without reducing into nothing.By nothing,i do not wish to avoke any enigma.We will try to peel down the layers of complexity,atleast to the point we can approach without mathematics.Most compelling explanations of the nature arises out of abstract concepts.Concepts like energy ,spacetime exist only in the realm of mathematics.It is better to try not to attribute too much meaning to them through our senses.Our intuition is the result of the mental models we carry about the world.Our limited capacities have rose to cater the needs of survival,not for deciphering the laws of nature.Our intuition of the world is based on our experience under small spectrum of the magnitudes.

Take for example Mass.whats the minimum mass range that we could tell the difference?May be we could tell the difference between a gram and some fraction of it not of lower orders and similarly the upper limit may be between 100 and 1000 kilograms(dont ask me to device an experiment for that..).

similarly we can establish limits to our perception of time and similarly length.But universe operates on far grander scale. we only have interaction with 7 to 8 orders of magnitude while the nature operates on the orders of 60 to 80 times greater–these are only the minimum estimates.

So its important for us not to confuse facilitation with the actual understanding and the ability to make predictions.Its a wonder how facilitation under such short range of operation allowed us to develop mathematics and probe rest of the universe.

3.1.3 Rocks and Brains: Why We Might Not Be Special ?

From now on we will take you on a journey of exploration.Here and there we would derail from the main theme to explain certain scientific ideas that would help appreciate the material more deeply.

Here we would try to impress upon the reader how complexity(we will try to define it in later sections,hold on to your intuition till then..) can emerge out of simplicity under suitable initial conditions when alloted sufficient time.We will understand evolution in all its grandeur.We will start by seeing how complex phenomenon can be explained by successively simpler rules.

If anything in science that most shocks me than it is that we are mere survival machines geared towards successful transfer of genes.The Evolution by natural selection is not the most bizzare of ideas in science but it is the one that has attracted most contempt.Evolution of complexity of sorts not just biological doesnt require any intervention– no fine tuning of the laws,no vigilence of any intelligence.All complexity is manifestation of

chains of simplicty.All processes are the play of atoms aimlessly roaming.The fate of these atoms which seem to be dependent on the precise attractive force between the neucleus and the electrons (this force being very gentle is the cause of all the complexity of the universe , unlike forces within nucleus which are much stronger but shorter in range these forces are perfect for the occurance of reactions or the transformations that eventually lead to us)are infact modulated by a tendency towards chaos.

> Francis Crick: "The origin of life appears to be almost a miracle, so many are the conditions which would have had to be satisfied to get it going."

The forces inside the atom are of two kinds–those between the nucleus and the electrons and those that hold the nucleus together.The former force is governed by a constant(approximately equal to 1/137) that determines the size of an atom.This has far reaching consequences.If that constant had been twice as big we would be half our size and have taken 10^{10} years to evolve.(Age of our universe is years).Forces within the nucleus must be very strong as it is very densely packed–If a human were a nucleus electrons would be roaming 100kms away.These forces are again only operational within the nucleus–If they could also influence electrons then there wouldnt be any reactions and thus all change halts.

In all the processes all that we see is chemical reactions–mysterious blobs of matter coming together under the settings of the right environment .These blobs are made up of atoms whose nature allows them to form close groups–molecules.When these molecules are set in warm liquids and allowed sufficient time–they tend to form bigger and more complex molecules.Under the pressure of dispersal,molecules stumbled upon reproduction.Some of them were able to produce their replicas under the conditions of the environment.Now change addition to neighbouring molecules added to the diversity of the produced children.Those that are more fit at replicating gave rise to apes and thinking machines.

But ,what drives these changes ? All changes are observed to increase the entropy of the universe.This refers to all forms of change from planetary motion to the formation of per-

ception.There seems to be a physical quantity that always increases with time.A slightly different phrasing might gives us a different expectation–all processes are directed towards increase of entropy.The second statement seems to indicate some cause that is increasing this physical quantity otherwise how can something increase out of nothing.This has raised entropy to an enigmatic stature–seemingly educated people making weird statements.

Infact it turns out that law of increasing entropy is not even a law–in that it imposes no constraints.Consider 10 coins arranged in a row.Now you are tasked to randomly select one of the coin and toss it.You are trying to keep track of the number of heads in the experiment.Now lets assume initial configuration started with all tails.Now as you proceed through the experiment on average system will evolve towards increasing number of heads.As all the events are purely random,lets assume on trail 1 you got heads.Now we have 9 tails and one heads.In the next trail you are more likely to choose to toss one of the tails than the only head.Irrespective of the outcome of this trail–head or tail ,total number of heads in the system would either increase or remain same.

Now consider the situation when you have 5 heads and 5 tails.Any further trail ,on average cannot change the count.When the number of coins increases–to the proportion of number of gas molecules in a room this nature will become increasingly evident.Irrespective of the initial configuration,system will evolve to ultimate chaos when not restricted.

Energy–an abstract concept that we associate with the capacity to do work varies in quality.Energy can be in various forms–from gravitational to thermal.When you take a room full of gas to describe its state(i.e the position and velocity of each molecule) we need vast amount of information.So we contend ourselves with average values like temperature.When we measure temperature of an object we have very little information of their actual state.Whereas consider a body raised above ground.The gravitational energy associated with it is same for all molecules.This is said to be of high quality.This tendency of energy to move from high quality to low quality–is captured in the Second Law of Thermodynamics or the law of increasing entropy.

But we can see how non-intervention and letting the system behave randomly can give rise to a fundamental law of the universe.The nature of the universe is inherently purposeless and needs no explanation.

But,our experience with other physical laws tells us that they are after all some constraints in behaviour–from newtons laws of motion to Maxwells equations of electromagnetism.As our understanding is deepening we began to find much simpler explanations to these phenomenon.With the formulation of Einsteins General theory of Relativity and formation of Quantum mechanics we began to see extraordinary simplicity in the workings of nature.It seems universe operates under no constraints at all.We learned that particles are waves and gravity is nothing but warped space-time.(2)Our notions of space and time reduced to one,so are of particle and wave.

Lets return back to entropy.Universe is drowning in chaos.Emergence of local order must inevitably creates chaos somewhere.Then what is stopping the collapse of energy–keeping it from flowing too fast.When we look around us there seems to be no where else in the universe that a life form controls the direction of flow of energy except on our planet.Rest everywhere universe seems to be mindlessly burning energy but it is all backed by a motiveless disintegration into chaos.Life and the apparent orderliness is afte-rall a brief trap before energy falls down the chain into more chaos.Gravitational energy has the highest quality.It protects us from the collapse.Freeman Dyson–mathematician and physicist in Energy in the universe talks about other hangups that slows down this flow of energy.Other wise we should have been incinerated by now.Its likely that ours is the only universe that survived to ponder about its existence.

Even if we reduce all the complexity to a few basic phenomenon even them we seem to be left out with the task of precisely defining some forces and some fundamental particles whatever they turn out to be.

Things behave and rules are our commentary on their behaviour.

Now we have established all change as motiveless degradation of quality of energy.Now

we will displace all the laws of universe.Things have their innate nature and their behaviour is in accord with their nature.We will see how laws are mere emergent illusion when the fundamental particles are let loose without any restrictions.Our perception and ultimately all our investigation is enabled by the fact that we can sense the world–see and feel.It all starts with light.We would Observe its behaviour and try to infer its nature.Than we can discard all rules.

We are taught light travels in straight lines and that they only bend at the interface of the separating media.This seems to be the rule that light follows,which is different from how other waves like sound travel.Its also observed that speed of light is controlled by the medium of travel .Now, as it turned out this makes light bend at the interface of the media.Why does it do so ?

A simple idea explains it elegantly–of all possible routes ,light travels in a path that takes least time.Similar rule–law of least action can be attributed to the motion of particles under force.

Let us consider two arbitrary points(a ,b)on the either sides of two media A and B stacked on top of each other.If a light ray from a has to reach b it can choose to cover some distance in media A and some distance in media B.We can imagine some optimum values for these distances that ensures the minimum time requirement of our rule.Snells law captures this requirement by specifying the bend of the ray at the interface based on the refractive index of the media.

But,why does light hurry so much ?And how does it know in advance the briefest path ? In a perfect world,there should be no rules, just matter and its corresponding nature should be sufficient(at least until we dispel that too).Our imposition of rules are the manifestation of the nature of matter.Light is a wave so it behaves so.Thats all we need to account for all its observed behaviour.

Imagine if we were to let light travel in all possible ways.A wave is a periodic undulation

of peaks and troughs.Imagine a light wave travelling from A to B.This will take any path unless forbidden.So it will reach B,through all possible paths.At B the resultant is the addition of all light arriving from all possible paths.So some will be crests,some troughs and a vast majority in between.But for every path there is a counter path that has opposite phase thus annihilating the effect.This means light cant travel at all.But ,consider straight line from A to B.This seems to be the only path surviving interference effects.Other waves that can bend like sound and radio waves have more wavelength to survive slightly bent waves interfering.The fact that these are waves is enough to explain its phenomenon.

Lifes complexity can be traced down to genome.In previous parts i have tried to give a general flavour of how almost no rules or at the very least from utter simplicity–exceedingly complex phenomenon can emerge.In biology the abstraction of the genome offers perfect vantage point of study.To create the right architectureas we have convinced ourselves that our brains are the inspiration.Consider your are an alien species visiting earth found a watch on the rock and trying to understand the working of it.Now,they hammered it to find out–they will see nothing but springs and wheels.Now we need to understand how each of them come together towards the working whole.The difficulty of the problem arises out of this complexity of interactions that we have to understand to come up with our version of the watch.

Our instruction set is written in DNA.Its long chain of four neucleotide bases–A,C,G,T.When you spell all these letters they constitute 3 billion letters for humans.These letters constitute the plan of a human.Now suppose you have a plan for a building.It includes all the information about everything in the building–the bricks,mortar,electrical system,plumbing system etc.Now suppose you need to fix plumbing system for one particular room in the building.The plumber would take a copy of the plan–but only those pages that describe plumbing of that room.Now biology does very similar–here RNA polymerase reads the DNA ,finds the code for a particular protein and transcripts.This transcript is called–mRNA.Proteins are what we are made up of.But how does these pro-

cesses represent life ?If we do not confuse life with consciousness then its pretty simple.Life metabolizes(uses up energy) and replicates.But how can we define life as it emerges from inert matter–as i warned take away consciousness(we will get to it later)and think what constitutes life ? The fundamental character is that its structures can edit themselves–the letters of the genome can mutate and change sometimes in response to the environment.Unlike machines that are engineered and behave in predictable fashion the very definition of life is evolution and unpredictability.Now we want to understand why does the phenomenon of life even exist ?

Now lets briefly look at computers–our greatest creation.Back in 1930s Alan Turing realized that there is a way of building a machine that can in principle perform any logic–he called Turing Machine.He deviced a minimal mathematical description of such a machine. After few decades John Von Neumann decided to build one such machine.At that time they used vacuum tubes and valves for the construction.They typically have to solve the problem of making a machine that always gives right answers based on fundamentally unreliable circuitary.We still use the same architecture that they deviced in 1950s but only instantiated on different hardware.The mathematical building blocks remained the same.Now our problem is the inverse–we are given life and asked to reverse engineer it.Our objective in deciphering the brain atleast for A.G.I is in understanding its computational principles–that might require us to understand all the biological detail.

3.1.4 Why Our Brains Are Not Optimal designs ?

Before we embark on this process its important to understand few details of evolution.Evolution for that matter any change is a purposeless process–our brains are not optimal designs.As the Nobel Laureate Francois Jacob (1977) put it, evolution is a tinkerer, who often without knowing what he is going to produce uses what ever he finds around him, old cardboards, pieces of strings, fragments of wood or metal, to make some kind of workable object [The result is] a patchwork of odd sets pieced together when and

where the opportunity arose.In the same token our brains gullibility(Kahneman, 2003; Tversky Kahneman, 1974 or refer Thinking Fast and Slow) clearly indicates our brains are merely a workable object , a compromise forced by the limitation of resources(an infinitely rational or a machine that could make optimal inferences in all situations would consume infinite resources.) and the constraints of evolution(Evolutionary inertia (Marcus, 2008)–the idea that evolution is constrained by previous history).

When evolution figures out better solution older ones are not discarded.This is the reason we still possess our old reptilian instincts–they take over us in times of pressure.We struggle between reflexive and deliberative(Marcus, 2008) modes even after realizing the merit of one over the other.

> Gary Marcus: "One of the saddest ironies of human nature is that we are simultaneously clever enough to make thoughtful, long-term plans yet foolish enough to abandon those plans in the face of temptation–while still being intelligent enough to feel remorse about it. Such tensions may not persist because they are intrinsically adaptive but simply because evolution, lacking forethought, could not figure out how to do better."

Its important to not make unnecessary assumptions in our search for AGI.That would seriously impede our development.Early sybolists thought mere juggling with symbolic representations would solve A.I–largely driven by a very wrong assumption which later Connectionists came with neural networks that allow for forming much richer representations.Tiny developments in the right direction are far better than hastening into dead ends.

From this section of the text i would take a bit formal stance,for those of you who have come here for some artificial intelligence stuff–here it begins

In a way–field of Artificial intelligence has become much easier.Neural networks have become lingua-franca in all possible domains–from computer vision to complex control

problems.A more representative word would be universal function approximators. Their limitation is evident–they would be great for narrow intelligence.Only reason of their recent success is beceause of the abundance of data and computing power.So,unfortunately both of them are going to improve in coming years which not only fuels their performance but also corresponding research and funding.It is important to realize why we started all of this in the first place.

Traditional pattern recognition problem deals with finding optimum or valuable features of the problem.For a long time people have tried to use logic based and purely mathematical methods in hope of one day coming up with human-level A.I.This has halted the true development of A.I for decades.

Alan Turing has proved mathematically that all computing devices are essentially equivalent and in proving that he imagined a 1. processing box,2.a paper tape,3.a method to read and write off it.He called it–Universal Turing Machine.In later decades this has set off computer era.Though he couldnt figure out how to create human level intelligence,proposed a test for finding if we have reached the goal ,the famous Turing test when an A.I converses with a human and tricks him into thinking that its a human.This focus on behaviour rather than the underlying process has made people to think that brain is a symbol manipulating device.Language is manipulation on words and vision is manipulation of objects and their locations.So,it doesnt matter how you implement it–as turing proved.If brain uses network of neurons ,we will use transistors.

In 1943 neurophysiologist Warren McCulloch and the mathematician Walter Pitts published an influential paper where they argued how neurons can act as logic gates.This has essentially proved that brains contain the fundamental building block of the computers.So without further biological proof they assumed that brains are nothing but computers.In later decades A.I pundits even predicted that they are close to human level A.I (its funny how similar thing has happened with physics before the discovery of quantum mechanics and relativity).

Its important to realize how associating intelligence to outer behaviour is not the right method to access our development.It makes no sense to say that Deep Blue–the famous chess program that beat Gary kasparov has any understanding of chess.Turing Test only makes sense when we have reached the destination.Meanwhile we should try to emulate human brain as almost all the state of art developments in A.I have come from some form of inspiration from neuroscience.[6]

Creating human level A.I is hard.Really hard.The search space of possible solutions are vast and we can easily be tricked.Our only source of inspiration is our brains.I do not say there cant be other ways to intelligence but only that searching for them wouldnt account for a good stratagy.Even neuroscientists seems to not have much idea of how intelligent processes happen within brain.Much of our knowledge about brains came from Functional imaging.However they cannot capture rapid changes.So,not until recently we couldnt even look at the processes happening in realtime.For long time the field of A.I has fancied with creating interesting things–hand writing recognition,anomoly detection and other pattern recognition problems.These are of great practical value but this has got little to do with our understanding of intelligence.

Its also important to distinguish superfluous details from truly important ones.Our brains are filled with lot of garbage–evolutionary junk.I suppose our solution to intelligence would be more elegant than our brains as they have evolved to cater very different needs.

3.1.5 The Evolution Of Brain

Around 3.5 billion years ago multicullular organisms evolved.They offered better rate of survival.With them the complexity of their processes have also increased.The nervous system allows for fast responses via electrochemical signaling and for slow responses by acting on the endocrine system.Nervous system accounts for the (1)evolution optimal sensors and effectors that allow it to maximize its control given finite resources and (2) evolution of a behavioral repertoire that maximizes the information gained from the environment and

generates optimal actions based on available sensory information.

So our brains primary motive was internal coordination–allostasis(refers to the predictive regulation of biological parameters in order to prevent deviations rather than correcting them post hoc).Our sense of joy on the achievement of goals,feelings of love and sorrow that direct our actionsintrinsic rewards–are evolutionary hack to improve our survival rate.Our capabilites to go through short term difficulties(inherently unpleasurable actions) are our learned association with them.So our behaviour arises out of this continual competition between drives and incentives that have adaptive value[4].

We have spent most of our evolutionary time as hunter-gatherers which have influenced our behaviour patterns.Our goal is to understand the brain at algorithmic level not implementation level.All the complexity attributed to the variety of cells and the wiring of neurons in the brain when seen in the light of understanding the basic algorithm would seem a lot less daunting.More importantly we should focus on the neocortex(the seat of intelligence where much of language and reasoning happens) and its remarkable homogenity of structure.In 1978 Mountcastle published a paper titled An Organizing Principle for Cerebral Function. In this paper,he points out that the neocortex is remarkably uniform in appearance and structure. The regions of cortex that handle auditory input look like the regions that handle touch, which look like the regions that control muscles, which look like Brocas language area, which look like practically every other region of the cortex.

Chapter 4

The Conclusion

4.1 So,what's wrong with the current approach ?

Consider what according to me is the coolest achievement of the present state of art(march,2018) in A.I–we have made algorithms(AlphaGo) that beat human expert in the game of Go(considered more subtle and hard game than chess for machines).AlphaGo is initially trained on 28.4 million positions and moves from 160,000 unique games(by using a combination of deep convolutional neural networks (convnets) and Monte Carlo Tree search played by human experts; it then improves through reinforcement learning, playing 30 million more games against itself. Between the publication of Silver and before facing world champion Lee Sedol, AlphaGo was iteratively retrained several times in this way; the basic system always learned from 30 million games, but it played against successively stronger versions of itself, effectively learning from 100 million or more games altogether. In contrast, Lee has probably played around 50,000 games in his entire life.

Maybe thats not an accurate way to compare given the pretraining humans receive in terms of the transferable skills they bring to the table even before they start learning anything unlike neural network based models that are trained from scratch.But there is a fundamental differences in the flexibility of our learned representations Consider various possible variants of the game Go.The wikipedia page Go variants describes versions such as playing on bigger or smaller board sizes (ranging from 99 to 3838, not just the usual 1919

board), or playing on boards of different shapes and connectivity structures (rectangles, triangles, hexagons, even a map of the English city Milton Keynes). The board can be a torus, a mobius strip, a cube or a diamond lattice in three dimensions. Holes can be cut in the board, in regular or irregular ways. The rules can be adapted to what is known as First Capture Go (the first player to capture a stone wins), NoGo (the player who avoids capturing any enemy stones longer wins) or Time Is Money Go (players begin with a fixed amount of time and at the end of the game, the number of seconds remaining on each players clock is added to their score).Human players can understand these variants and adapt to them because they explicitly represent Go as a game, with a goal to beat an adversary who is playing to achieve the same goal they are, governed by rules about how stones can be placed on a board and how board positions are scored. Humans represent their strategies as a response to these constraints, such that if the game changes, they can begin to adjust their strategies accordingly.

4.1.1 The Future Ahead

Nearly four centuries ago,telescope was invented–we learned more about the universe in the next decade than in all human history before.FMRI scans and other non-invasive technology is growing at exponential pace.This would inevitably continue.

Not until recently we werent able to simulate models of the brain that are complex enough to set up actual requirements.A huge bottle neck has been the availability of data and computational resources.The future is exciting

4.1.2 Law Of Accelerating Returns

Ray kurzweil a famous inventor observed a peculiar trend in information technology that he divided into six epochs.As information is passed to further generation the developments are hastened thus growth is exponential.He predicts somewhere in 2040s as the singularity

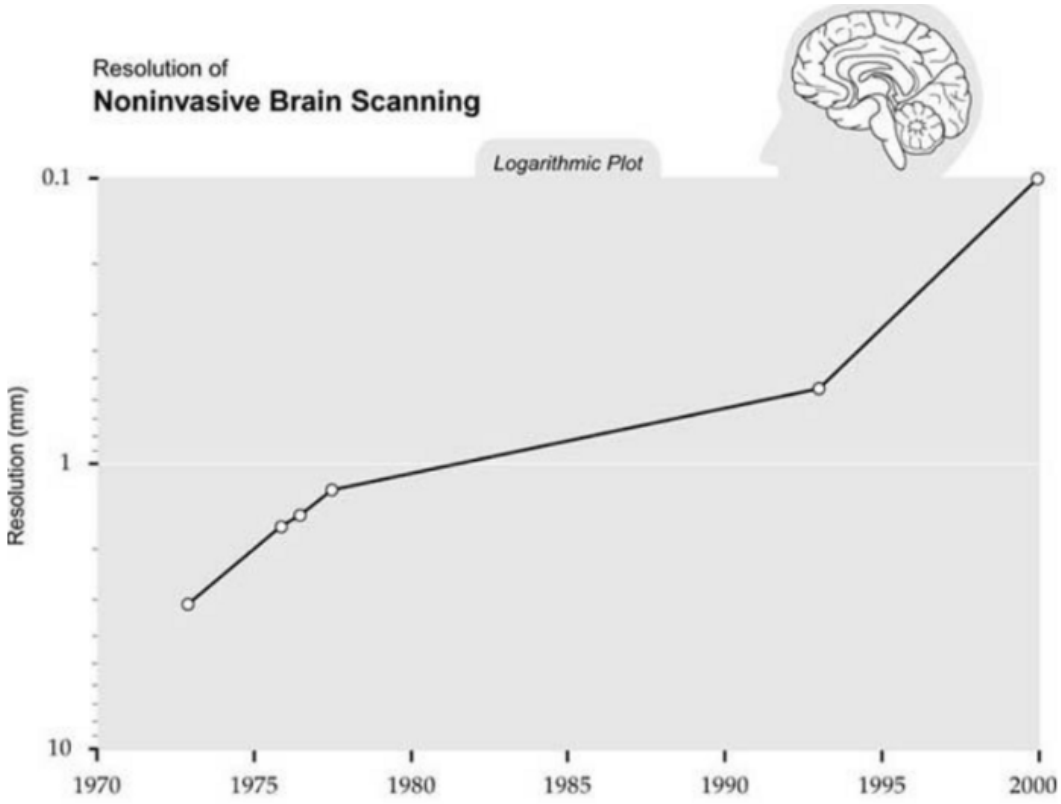


Figure 4.1: Brain Scanning Technology[7]

point where we merge with technology and ultimately greater than human intelligence drives the development.In The Sigularity Is Near he defines singularity as the point in human history where rate of development is so fast that it seems nearly infinite and eventually our intelligence would spread the universe.I would not take up the task of predicting the time but if we manage to survive along the line than Singularity as kurzweil defines is inevitable.

Physicist and complexity theorist Theodore Modis analyzed clusters of events (which he calls canonical events)that resulted in increase of complexity and order,which when plotted on logarithmic scale shows clear trend:

> Ray Kurzweil: "A billion years ago, not much happened over the course of even one million years. But a quarter-million years ago epochal events such as the evolution of our species occurred in time frames of just one hundred thousand years. In technology, if we go back fifty thousand years, not much

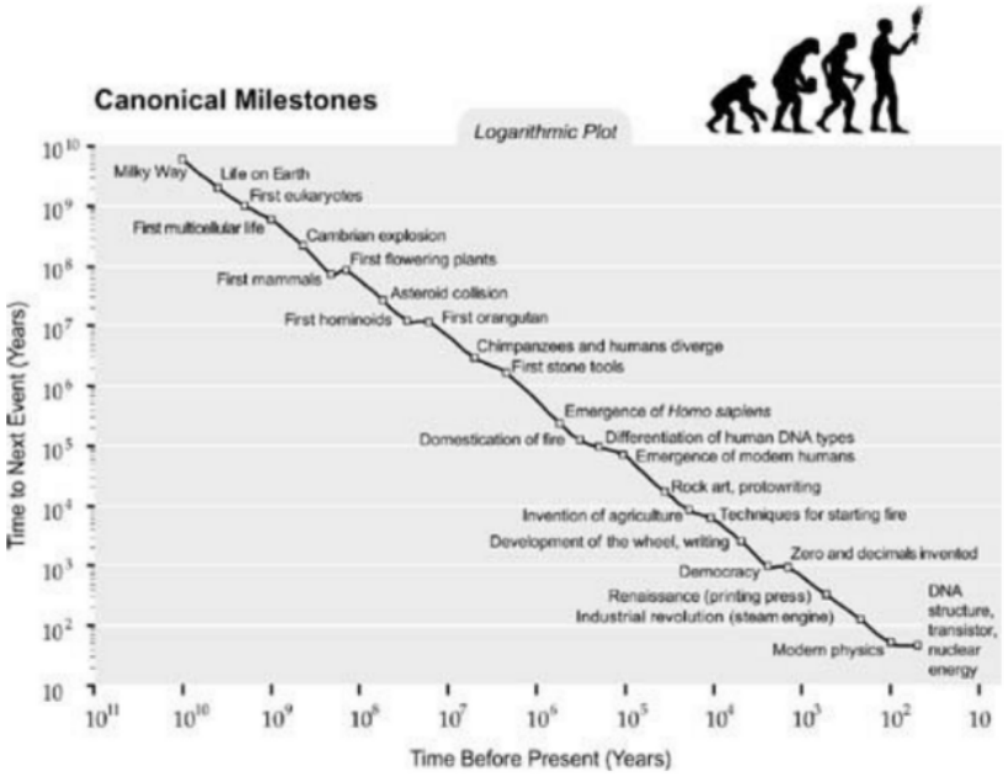


Figure 4.2: Exponential trend in the increase of order and complexity.[7]

happened over a one-thousand-year period. But in the recent past, we see new paradigms, such as the World Wide Web, progress from inception to mass adoption (meaning that they are used by a quarter of the population in advanced countries) within only a decade."

The essential argument is Moores law which is attributed to electronics is part of a more general trend of information technologies where future growth is dependent on the present state.In case of computing power the rate of growth itself is growing as the slope itself is increasing.This would bring us with personal computers that outperform our brains in the next four years.(brain is estimated to operate at 10 flops whereas present supercomputer can rack up to 10 flops)

Imagine the innovations that would foster with the ability functionally simulate human brain.AGI is closer than expected.

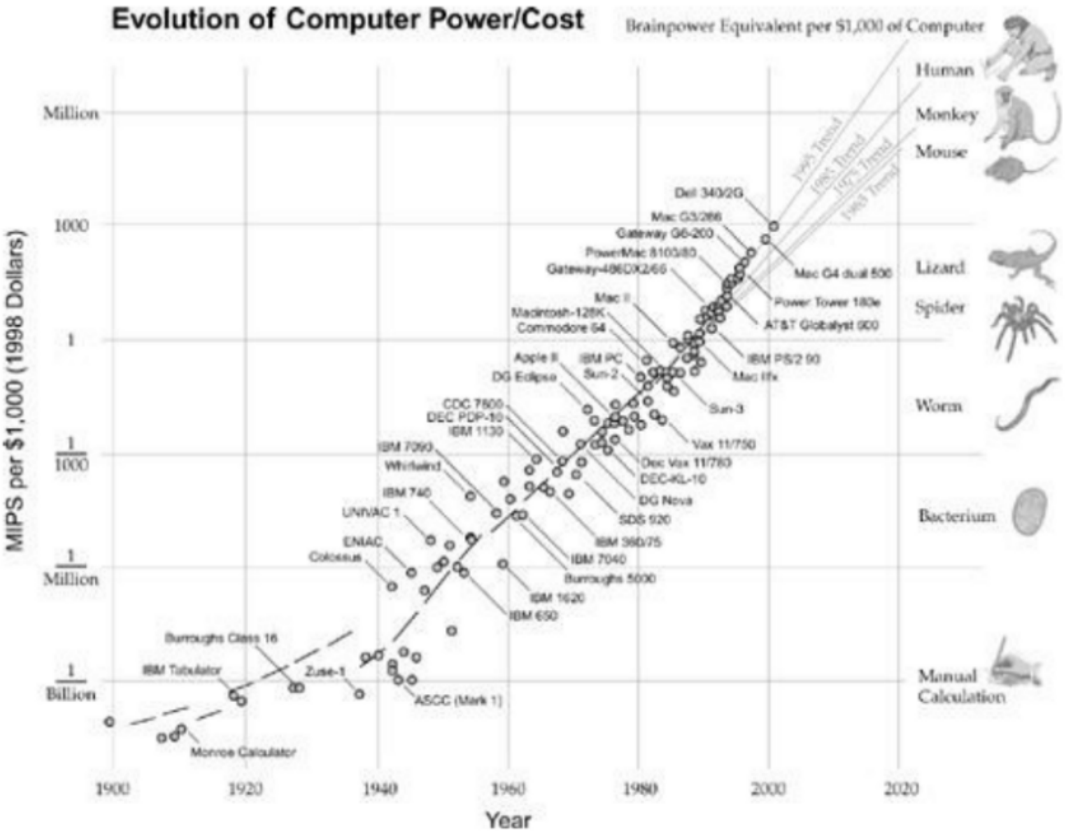


Figure 4.3: Compute power vs cost[7]

We wanted the reader to see that emergence of phenomenon as weird as consciousness can arise within the laws of physics,so in principle can be understood and reverse-engineered.The inevitability of A.G.I can only be contradicted if one assumes the need of some esoteric laws that are beyond the reach of physics.The varying phenomenon of cosmos are amptly being explained by very simple laws.

Bibliography

Articles

[1] A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay.US Naval Research Laboratory Technical Report 5510-026 arXiv:1803.09820 [cs.LG]

[2] Leslie N. Smith,Cyclical Learning Rates for Training Neural Networks ,*arXiv:1506.01186 [cs.CV]*

[3] Alex Krizhevsky Ilya Sutskever,Geoffrey E. Hinton,ImageNet Classification with Deep Convolutional Neural Networks,*arXiv:1606.01186 [cs.CV]*

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun,Deep Residual Learning for Image Recognition,*arXiv:1512.03385 [cs.CV]*

[5] Aurlien Gron,*Hands-On Machine Learning with Scikit-Learn and TensorFlow Concepts,pg no :13-100*

[6] Tom Everitt , Ben Goertzel Alexey Potapov,*Artificial General Intelligence,pg no:120-158*

[7] Ray Kurzweil,*The singularity is near,pg no:100-156*