Deep-Physics: Intuitive Physics of Dynamic Environments using DNNs, GANs, and RNNs

James Kary and Alexandre Lavoie

March 2019

Abstract

Despite revolutionary advancements in Artificial Intelligence, General AI remains a stagnant field of research. Even if Neural Networks have mastered a multitude of individual tasks, current models remain weak solutions of mimicking actual human learning. This research attempts to break away from Narrow AI by focusing on recreating the innate human perception and understanding of the world. More specifically, this research explores the concept of intuitive physics by proposing Deep-Physics. Using numerical data from physics models, Deep-Physics is trained to replicate human's intuitive understanding of physics. Given initial conditions of objects, Deep-Physics generates a realistic step-by-step approximation of dynamics of environments. The simplicity of the model allows Deep-Physics to be a potential alternative to real-time physics engines. Using a Deep-Physics model could be a step into a more efficient and intuitive Machine Learning, eventually leading to the ultimate goal of General AI.

1 Introduction

This section will briefly introduce Artificial Intelligence. We will discuss the current achievements and downfalls of Artificial Intelligence. In addition, we will explore new research into a more universal, General Artificial Intelligence.

1.1 Current State of Artificial Intelligence

Artificial Intelligence has had a major breakthrough during the last decade. New hardware and techniques allowed Machine Learning, a major contributor to the Artificial Intelligence movement, to flourish. The vulgarization of the mathematics, through a variety of libraries, has made Artificial Intelligence mainstream. Machine Learning can now be found in virtually every scientific field and multi-billion dollar corporations.

Most universities have teams conducting research aimed at the innovation of artificial intelligence. Machine Learning has become one of the most popular methods to solve scientific problems in diverse fields. Biologists use it to try to understand protein synthesis, Astronomers use it to model merging galaxies, and Computer Scientists expand on the large Machine Learning library. There is no denying that Artificial Intelligence is contributing greatly in academic advancements.

Companies are putting large incentives towards Artificial Intelligence. Google Brain, Open AI, Facebook AI, of many, are heavily funded branches of very large corporations. These teams are behind revolutionary projects like Alpha Go and Gym. In addition, research from these departments help e-commerce, social medias, stocks, just to name a few applications. Thus, Artificial Intelligence is a major player in today's economy.

Considering the popularity of Artificial Intelligence, why does it remain so limited? Why are we training Machine Learning models that can solve singular tasks? The problem we are facing is that our Artificial Intelligence is still "weak". More research is currently being conducted in hopes of creating an Artificial Intelligence that compares to Human Intelligence.

1.2 Research into General Artificial Intelligence

General Artificial Intelligence is still a quite recent field of research. The directions that the research should takes remains ambiguous. How can we make our "Weak AI" into "General AI"? One revolutionary paper believes that we should change our perspective on Machine Learning.

(Lake, Ullman, Tenenbaum, & Gershman, 2017) propose that for an Artificial Intelligence to be more "General", it should more closely mimic Human Intelligence. Our current Artificial Intelligence models learn from scratch, which this paper believes is it's weakness. There are a variety of "innate" intelligence and skills in newborns. The paper contains an extensive list of them, one being "Intuitive Physics". The research hypothesizes that an Artificial Intelligence should possess this fundamental intelligence to achieve Human Intelligence.

In an attempt to prove this hypothesis, our research examines further Human Intelligence. We more specifically focus on Intuitive Physics. We examine how this knowledge is present in children and apply it to Machine Learning models. In combination with other "innate" intelligence from (Lake et al., 2017), we hope that this will be a step towards General Artificial Intelligence.

2 Human Intelligence

Artificial intelligence at its core has its foundations based on a fundamental idea of intelligence derived from that of humans. This section will explore important psychological aspects of intelligence that will serve as a basis for the later discussion of cognitive processes and evidence regarding the shortening of the gap between Machine Learning algorithms and the function of the human brain.

2.1 Intelligence as a Multidimensional Idea

What does it actually mean when an individual claims to be smart or intelligent? We all have an idea of what they mean but it remains unclear what the concrete meaning of it entails. The truth is that there is no one definition and its meaning depends on the context in which it is asked. The definitions most people have of intelligence go to the extent of intelligence as being smart or intelligence as what one does when they don't know what to do. We often refer to it as a persons cognitive abilities such as; learning, memory, and attention, yet this leaves a lot to be described. Misconceptions about intelligence are perpetuated by the conflicting information provided by psychologists of different specialization. Let's begin by looking upon an individual widely considered to have high intelligence; Derek Pavacinni. Derek can play any piece of music on the piano while only having heard it once and can play it in any musical style. He pleases crowds with his talent, capturing amazement from large audiences. Derek is what is referred to as a savant; an individual gifted with an extraordinary mental ability. Derek has low IQ and and is unable to care for himself due to being born blind. Most savants have low IQ and cannot care for themselves, however, they are important examples of the variety of cognitive abilities that exists. The question of why we can't have the same mental ability as Derek raises an important point regarding the definition of intelligence; the uniqueness of a mental ability does not resolve the remaining aspects of intelligence and thus cannot be considered as such. Noting that Isaac Newton and Albert Einstein did not bear the same memory abilities as Derek, yet they are two of the most intelligent minds in history, we begin to uncover the fractal nature of intelligence. Psychologist Haier (2017) elaborates by taking the IBM Watson computer as an example. It was able to beat the two Jeopardy world champions using its 15 petabytes of training data. Does this mean that it is intelligent? Is Watson more intelligent than Einstein for the simple reason that the latter didn't accomplish the same task? Taking a closer look at some conceptual definitions will perhaps aid in reconciling these questions. Gottfredson of the University of Delaware provides insight as to how the concept of intelligence should be depicted in :

> Intelligence is a very general mental capacity that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test taking smarts. Rather it reflects a broader and deeper capability for comprehending our surroundings—"catching on," "making sense" of things, or "figuring out" what to do. (Gottfredson, 1997)

Where does that leave the "intelligence" in "artificial intelligence"? Surely both definitions cannot be referring to the same concept. More will be illuminated upon in 4.1, but for now let's explore in detail why it is of interest to lead the field of artificial intelligence in the particular direction of motion prediction.

3 Intuitive Physics

This section will discuss evidence and the cognitive processes involved in what is called the "Physics Engine in Our Mind". In exploring various studies conducted by well-known computer scientists, it will be possible to understand the significance of creating a Machine Learning algorithm that has the ability to predict the physics of a specific scene or object. Starting from earliest research to the state of the art models of human intuitive physics.

3.1 Historical Note

As humans, our perceptual understanding of the world is largely understood through the intuitions we draw about the behaviour of objects around us. What perhaps is most surprising is how we begin to draw these inferences at a very early age of infancy. Developmental psychologist Jean Piaget is widely known for establishing the traditional view of how infants perceive the world. In his research, he concludes that all humans are born with innate reflex-like sensory motor schemas that allow them to perform very simple behaviours. Through bootstrapping of these schemasand aquring experience, more complex behaviours are produced. It is then that the infant becomes conscious of the distinction between self and other and self and objects around them. This view, despite being once popular, has been redefined in modern psychology from the advent of innovative research techniques aimed at understanding the cognitive abilities of infants.

Of these research techniques, the dishabituation paradigm is arguably most influential as it has made it possible to better understand and manipulate infant cognition. This paradigm provides a basis for measuring cognitive faculties of infants, in which time spent looking at an event serves as an indicator of degree of surprise. Developmental psychologist Baillargeon (1986) showed that infants have certain expectations about how objects interact in her drawbridge experiment. By first habituating the infant to the motion of a drawbridge, a block outside of their view was placed as to obstruct its motion. Baillargeon observed that infants looked a lot longer when the drawbridge's motion was obstructed. She was then able to conclude that they show a greater degree of surprise when faced with an event that involves a certain violation of expectation. Furthermore, her findings fueled new research avenues, branching off from the traditional view, as her experiment demonstrated how infants of much younger age than previously thought are aware that objects exist despite not being directly observed. Experiments such as this one and others include within them the the seed of what is referred to as "naïve physics" or intuitive physics.

Intuitive physics then captured the attention of Spelke (1993), pioneer of experiments involving the dishabituation paradigm. She firmly defended her stance in which human infants possess the cognitive capability to parse their visually perceived universe into finite individual objects which abide in accordance to principles afforded by physical reasoning. Her research focused on four aspects of the intuitive physics observed in infants. Together along with an understanding of it as a theory built upon laws and principles, it serves as the building blocks for the computational model of infant naive physics. The principle of cohesion illustrates infant view of objects as ocurring together only if such surfaces occur together. To understand the perception of infants faced with the motion of objects, the principle of contact is used to depict how perceived objects moving together must in turn be in contact with one another. In addition, she conjectured that there must be some principles that allow them to distinguish distinctive properties of objects such that object motion through space and time is restricted resitricting it to having only one position at a given time. Spelke (1988, p. 161) describes infant intuitive physical inference:

> I suggest that the infant's mechanism for apprehending objects is a mechanism of thought: an initial theory of the physical world whose four principles jointly define an initial object concept.

These governing principles of reasoning are those of solidity and continuity, which have application in the aformentioned drawbridge experiment conducted by Baillargeon (1987). The drawbridge experiment showed how infants are sensitive to their being more than a single object at a certain place at a particular time, which exemplifies the solidity constraint. Spelke demonstrated how infants make use of these reasoning principles to make physical inferences about objects in their environment.

The inferences made by infants about unobserved phenomena using previously perceived concepts is analogous to scientific procedures. It is as though they are all little scientists trying to explain the behavior of objects using their sense of vision to discriminate between objects and infer as to how they will likely behave. These inferences become the foundation of what will become the storage of theoretic physical knowledge. From an information processing perspective, these inferences are represented symbolically from which the infant computes a prediction regarding the motion of an object. This sometimes results in the infants surprise at the contradiction between computed and perceived results in a dishabituation experiment.

Computational models and theoretical models of infant intuitive physical reasoning are flawed in that they rely on the assumption in which mechanisms interact with principles to produce behaviours. They operate under the strong assumptions that the cognitive system interacts with principles of physical reasoning in explicit symbolic representations. This implies that the infants access these principles to reason with their surroundings as to evoke certain behaviours. Munakat, Mclelland and other connectionist modelers built the first neural network model of infant naive physics. They showed the computational and theoretical models wrong by making a neural network act in accordance to the continuity principle without it being explicitly represented in symbolic form. Their model reflects how infant knowledge of object permanence is stored implicitly in patterns of neural activation that vary in strength and firing rate, with the use of pattern recognizing associative mechanisms. Intuitive physics in infants was shown to be a matter of persistence of specific neural activation patterns of object representations rather than a set of fixed coded principles, as suggested by Spelke.

The implicit understanding infants have of object physical behaviour is the foundation of what becomes an extensive theoretical understanding of physics at later stages of life. Beyond infancy, this understanding shows itself in the many intuitions about the behaviour of physical objects when interacting with the diverse aspects of quotidian life. Researcher's thus tasked themselves to understand how we come to form these intuitive theories of the physical aspects of our environment. The next section goes over research in cognitive neuroscience in search for explanations as to what cognitive processes underlie our physical inference capabilities. Is innate cognitive architecture responsible for the development of early physical reasoning, or is it attributed to the experience that fills the pages of the blank slate of an infant's mind?

3.2 Controversy in Intuitive Physics

According to what principles are human physical judgments constructed? In this section are highlighted the most salient controversial early research studies that address this question.

Research conducted by McCloskey (1980) suggested that people with no formal physics training employ principles of fallacious Aristotelian physics principles in absolute dissonance with those of Newtonian mechanics. Experimental trials in which students were told to draw the direction of a ball's motion following its passing through tubes of various shapes showed that people use impetus theory to make predictions. Such principles were popular three centuries prior to those of Newton. They are based upon the belief that objects set in motion acquire an impetus which causes its acceleration, gradually dissipating until it stops. A particular misconception is the "straight down belief" in which individuals predict that the trajectory of an object dropped from a moving body falls straight down, as opposed to the normative parabolic projectile motion path. McCloskey attributed this fallacious belief to a perceptual illusion in which individuals predict the object's motion from the reference frame of the moving body. Research on the naive theories (McCloskey, 1983) that people employ to make physical predictions suggests that they arise as a result of illusory perceptual experiences which are used to shape the concept of physics around them, leading to misconceptions.

Cooke et al. (1994) then argued that the naive impetus theory that McCloskey supported are of elevated complexity. It was discovered that they may not be the sole method of prediction employed. It was shown that individual physical judgments depend on more factors than simply the recollection of a naive theory of motion because only a marginal amount of participants used this. Cooke et al. propose that people construct theories of motion in an "on the fly" manner. They argue that people construct theories of motion using a combination of information from the contextual descriptors of the problem and the retrieval of theories or prior experiential knowledge relevant to it. Thus revealing that physical judgments are dependent are more factors than previously thought, exposing how the context and familiarity in which a problem is presented greatly influences the final judgment. Pivotal to the previous conclusions, it was shown that people's predictions accord with Newtonian principles more closely than when the problem presented is of a familiar context. Therefore separating the impetus theory and theories created "on the fly". This was further explored by Payne and Squibb (1990) which described the difference between a genuine error and a misconception as:

> generated by a desperate student who has no idea how to solve the problem. [p: 462]1990

Stirring greater controversy, the journey towards determining the mechanisms responsible for our ability to reason about physical situations began among cognitive scientists. Their primary concern is:

> With such extensive experience with moving bodies, it would be expected that one acquires a rudimentary understanding of the principles which govern moving bodies, yet misconceptions pervade without consequence, why is this so?

3.3 The Physics Engine in Our Mind

Moving on from infant physical reasoning, the next sections will address the "physics engine" in our mind that allows us to make perceptual judgments without the need for explicit Newtonian mechanic computations. What cognitive processes underlie our intuitive understanding of the physical constraints in our environment? By what means do we make such predictions and are they accurate when compared to Newtonian mechanics? Such questions have been the subject of various studies conducted by cognitive scientists, psychologists, and computer scientists. Our lives are pervasively filled with perceptual judgments about the behaviour of diverse physical events. From assessing the thickness of ice to predicting the motion of keys falling off a counter, our world continuously demands our judgment of its physical state at a future time. In this section we explore the predecessor models of human intuitive physics, the sources of uncertainty present within them, their accuracy, the assumptions made in using such models, and the noisy Newton framework.

3.4 Predecessor Models of Human Intuitive Physics

Here, we explore the earlier models of people's understanding of physics as well as the constraints and assumption made in using these models. In (Sanborn, Mansinghka, & Griffiths, 2013) these models are compared as too highlight each of their strengths and weaknesses in mass judgment experiments consisting of participants asked to predict the mass ratio between a motor ball and an initially stationary projectile ball.

3.4.1 Direct Perception

Gibson (1966) argues that our perception of the environment in which we exist is “direct” in that what we perceive remains unaltered by information from touch,smell, and other visual sentiments. As such, he suggests that perception is as its core based on the acquirement of information and not on having the sensation of visual stimulus. In (Sanborn et al., 2013), participants where asked to predict the mass ratio of a given collision between an initially stationary projectile object and a motile motor object. The model of direct perception in the context of the mass judgment carries with it the assumption in which people perceive the information pertaining to the final and initial velocities of the balls rather than gathering such information through predictions based on visual sensations. Therefore, it assumes that participants receive noiseless information regarding the presented velocities, allowing them to conduct accurate computations of the mass ratio between the two balls. This model claims that people’s perception of the object with the greater mass is independent of the ratio of initial and final velocities present in the collision. Due to the assumptions that this model imposes on participants judgments, this model is ineffective at providing a unifying approach to human perceptual judgments of physical constraints.

3.4.2 Kinematic Specification of Dynamics(KSD)

Runeson (1977) argues that motion points to the causal factors that underlies events. His approach of kinematic specification of dynamics focuses motion perception on the dynamic factors of an event rather than those of kinematics. Under this light the collisions observed in (Runeson, 1977) by participants can be parsed into their observed kinematic variables such as velocity and their dynamic variables such as mass. These variables together determine the path taken by the colliding balls. Despite not being able to directly predict the mass of colliding objects under this approach, Runeson showed that observing the causal factors of motion does enable the observer to predict the mass ratio between the objects colliding. This approach permitted Runeson to derive an equation independent of the established ratio of initial and final velocities. His equation suggested that people’s judgment of mass ratio should remain consistently accurate when subjecting the observed collisions to variances in velocity ratio. This evidently strong assumption was tested by Todd and Warren in (Todd & Warren Jr, 1982) where they conducted a similar study to that found in (Sanborn et al., 2013). They found that information relating to relative speed is favoured for mass judgments when the two objects are in motion before contact. Furthermore, they observed that when a motile object collided with an initially stationary one, the participants tended to base their judgments on other information given the properties of the collision. These where conducive towards erroneous responses when the relative velocity ratio between the balls is low, the mass difference is small, and when the stationary object is heavier.

Todd and Warren (1982) evidently found Runeson’s Kinematic specification of dynamics to be an inaccurate model of people's physical judgments due to the apparent deviations from Newtonian mechanics observed in people's responses. The erroneous responses observed elude to a systematic bias towards the motor object. In the instance where the motor object was heavier, people's responses were correct, however, when it was in fact the lighter of the two, people still judged it to be the heavier one. Therefore, this systematic "motor object" bias served as strong evidence against kinematic specification of dynamics and that people aren't endowed with the ability to judge object masses with any accuracy through simple observation. The next model attempts to consider these deviations in addition to providing an explanation to why they occur.

3.4.3 Heuristics

Pólya postulates that :

> Heuristic reasoning is reasoning not regarded as final and strict but as provisional and plausible only, whose purpose is to discover the solution to the present problem. (George, 1945)

Problems we encounter in our lifetime are generally believed to have correct solutions that are defined by means external to the process by which they are obtained (Myron L. Braunstein & (Auth.), 1976). Most often, solutions to such problems require accurate measurements of our perceived surroundings. Braunstein explains how the process by which we attempt to solve a perceptual problem in our external world generally lead to correct hypothesis relating to its solution. Perceptual processes that are used in problem solving and that lead to incorrect solutions are not typical in nature due to the forces of natural selection favouring species that make the least mistakes. Furthermore, he describes how some of these processes persist in that they regularly result in incorrect hypothesis pertaining to the solution of a problem. The class of procedures that are most often apprehended when faced with a problem is known as "heuristic methods". When solving a difficult problem in which not all probabilities may be known or the information required to reach a correct solution is unavailable, the most likely that of a heuristic process. We do this as to have the most efficient solution process, sacrificing the certainty of the correct answer, by imposing restrictions on the space searched for a solution. Therefore, heuristic processes, as described by Braunstein, are procedures that don't necessarily lead to the correct solution. They provide a method of restricting possible solutions found in the search space by analyzing the structure of the given problem leading to a conclusion.

Psychologists currently believe that the reason for the observed discrepancy between Newtonian mechanics and our intuitive understanding of physics presented in the preceding models is due to the various heuristics we involve ourselves in throughout life. In (Sanborn et al., 2013), the heuristics apprehended by the participants judging which colliding object had a larger mass consist of simple quantities such as: the lighter ball is the one that has the greater velocity or it is the one that ricochets. A heuristic model of human intuitive physical understanding that operates under the assumption that we are able to make judgments faithful to Newtonian mechanics in only very simple situations was then constructed. As a result, the collisions observed in the experiments conducted are largely understood in terms of heuristics. The suggested heuristic model provides a qualitative approach to Newtonian mechanics by including a function of salience for the chosen heuristic process in the participant's mass judgment. The model response is thus directed by the salient response of the participant with various conditions applied depending on the type and amount of salient heuristics used. The data is fitted a salience threshold which measures the degree of salience of the heuristic used. If there's only one salient response, then the model's response is as such. If the two above mentioned heuristics are salient then one is chosen to override the other. If neither heuristic is salient then the model responds by choosing a stochastic response of equal probability among the other possibilities.

This model seems to have promising application in the experiments present in (Sanborn et al., 2013), however, various parameters are left ambiguous in using this type of model in other situations. Type of heuristic used is significantly dependent on the situation in which the model is applied resulting in variability preventing this model's effective use in other situations. Therefore, this model of human intuitive physics fails to present a unifying account of how we truly perceive physics. The next section presents the framework which uses elements of all the previous extant models as to provide this unified account.

3.5 The Noisy Newton Model

Research in the cognitive processes involved in human intuitive physics has been reinvigorated in recent years with innovative models that have succeeded in mimicking our tacit physical judgments. Realizing that the degree of accuracy in people's physical judgments is positively influenced by people's familiarity with the task environment and negatively influenced by lack of background knowledge, a model was developed as to account for these sensory uncertainties.

Originating from psychophysical studies of behaviour aiming to measure individual reasoning of manifold dynamic scenes, the *noisy* Newton framework hypothesizes that human intuitive physics arises as a result. Research in mental simulation has shown that humans use cognitive processes which allow one to construct spatial representations of physical situation internally. This warrants the reasoning of possible dynamics present in the system through its manipulation. Furthermore, these internally conceived representations can be reasoned with despite being unbound to precise prior knowledge. This accords individuals with the ability to make judgments of the outcome of a dynamic display in which little may be known (Hegarty, 2004).

Research conducted at the anatomical level has shown that neural activity when performing physical inference tasks is primarily found in cortical regions of the brain. Such regions were found to be exclusively active during such tasks, however activation overlap between these regions and those involved in cognitive processes such as scene understanding was observed. This suggests that there exists a link between the cognitive processes involved in parsing the physical content in a spatial array and the regions of neural activation in the brain (Fischer, Mikhael, Tenenbaum, & Kanwisher, 2016). In (Sanborn et al., 2013) participants were asked to predict the mass ratio between a motor ball and a projectile ball, their judgments were used to train a probabilistic simulation model. A substantial number of collision event simulations were completed to effectively model human-level predictions. The coefficient of restitution, e, was changed independently from the mass ratio to observe how changes in the ratio of relative initial and final velocities affected people's judgments of the mass ratio between the colliding balls. Inputs consisting of uncertain sensory information were combined with ground truth physics principles of Newtonian mechanics. Perceptual and physical variables present in the physical situation are assumed to be inputs that are assimilated with noisy sensory information based on one's prior beliefs. These inputs are used to model the physical constraints of the situation in the form of expectations that follow Newtonian principles. Objective data from these inputs are converted into estimates subjective to each individual participant, accomplished by combining noisy sensory input with randomly selected prior beliefs of the perceptual cues in the situation (J. R. Kubricht, Holyoak, & Lu, 2017). Furthermore, mass are physical variables that were also inferred because such priors are dependent on ones knowledge and personal experience with the physical situation.

3.6 Modelling Using Probability Theory In AI

This section will aid in developing a more intuitive understanding of the probabilistic approaches to human physical scene understanding by giving a brief overview of the concepts underlying the implementation of probability theory in Machine Learning systems.

To fully understand why models developed using the noisy Newton hypothesis are influential in the field of human cognition, a brief overview of the basic idea behind Bayesian inference which these models utilize is presented.

Bayesian inference is the mathematical procedure in which prior knowledge is incorporated into statistical probabilities. It allows for one's subjective statistical beliefs to be updated when presented new information in support or conflict with two hypotheses.

The prior beliefs are denoted as the probability of "B" given "A" using conditional probability notation $P(B|A)$.

The updated prior beliefs are termed the posterior beliefs, denoted as $P(A|B)$ which is the probability of "A" given "B".

The posterior odds are computed using Bayes rule. For a variable θ and evidence y, we have:

$$P(\theta|y) \sim (y|\theta)P(\theta) \tag{1}$$

Here $P(y|\theta)$ is the new evidence presented for or against the hypothesis being tested, also known as the likelihood. P(θ) is known as the prior knowledge which are the prior beliefs before considering any evidence. The joint probabilities of data y given a parameter can be written as:

$$P(y, \theta) = P(\theta|y)(P(y)) \tag{2}$$

or equivalently,

$$P(y, \theta) = P(y|\theta)(P(\theta)) \tag{3}$$

Combining these, cancelling out P(y,θ) gives Bayes' rule for updating prior beliefs when presented evidence:

$$P(\theta|y) = \frac{P(y|\theta)(P(\theta))}{P(y)} \tag{4}$$

P(y) is the evidence. P(θ) is the prior beliefs. $P(y|\theta)$ is the conditional probability describing the likelihood, Likelihood describes the probability of a model generating perceived data. (Barber, 2011, p. 173)

(Tenenbaum, Kemp, Griffiths, & Goodman, 2011) provide an explanation for how humans are able to make significant inferences given very little information. It provides a rational framework aimed at demonstrating the guidance that abstract knowledge has on human learning and inference. Bayesian inference allows the discussed probabilistic simulation models of human intuitive physics to effectively model the performance of human predictions tasks involving the formation of mental simulations that allows for the inference of causal relations or the presence of possible latent variables. Below is an example of how Bayesian inference is used in a probabilistic simulation model.

Bayesian modelling is used to replicate human-level inference by assuming that one's specific hypothesis is limited by their past knowledge, imposing a foundational constraint on the total hypothesis search space for the problem. The judgment imposed on observed data is limited by one's background knowledge as the hypothesis space does not contain all the possible values of possibly observed latent variables. By using Bayes' theorem and iterative simulations, a generative model of human-level inference is developed by recursively updating beliefs about the latent variables in the situation.

Tenenbaum et al. (2011) presents an example of how this framework may be used in a general case:

> For an observer with hypothesis search space H consisting of values of latent variables leading to outcomes and is constrained by past knowledge. Let h be their specifically chosen hypothesis for the situation to yield a particular outcome given a set of latent variables.

The probability of the hypothesized outcome given the observed data, d, using Bayes' rule is:

$$P(h|d) = \frac{P(d|h) * P(h)}{\Sigma_{h' \epsilon H} P(d|h')P(h')} \tag{5}$$

In which the prior beliefs are denoted as P(h), beliefs prior to data observation. The likelihood indicates the probability of the observed data given the hypothesis, or how expected the observed data is relative to the hypothesis.

Thus, P(d—h') is the likelihood of all the other hypotheses and P(h') is the prior beliefs. Their product summed over all the h'ϵH is the summed joint probability of all other hypotheses being observed relative to the given observed data.

3.7 Probabilistic Simulation Models

It was then observed that the simulated physics found in a video game graphics engine and our physical understanding of the world we live in, have striking similarities. Both simulated physics in video games and the physics engine in our mind function as to represent physical phenomena using simplified accounts of Newtonian principles. On the one hand, our noisy sensory inputs generate inaccuracies in our inferences in physical judgment tasks, and a video game graphics engine functions with the use approximations of normative Newtonian principles to simulate believable yet inaccurately implemented physical constraints. Video game physics engines are analogous to that which allows us to infer on the outcome of a dynamic display as they both model the physical constraints in a scene by means of approximation of Newtonian principles. For the human endowed physics engine, uncertain information is combined with background knowledge of the observed variables which are then used to form a mental simulation of the physical constraints in the visual input. Such simulations are subject to error which renders them approximate. In this respect, the physics engine humans are endowed with corresponds to how a game physics engine simulates physics without explicit computation of Newtonian principles. It will implement constraints to the environment yielding approximated events guided by normative principles.

Models based on probabilistic simulation were then developed as to bring life to the noisy Newton hypothesis in Machine Learning. These models were shown to be able to replicate with accuracy the way we perceive and reason with physics. Research based on probabilistic simulation have tasked themselves with identifying a clear parallel between human-level physical judgments and those generated by such models. These models replicate our ability to form internal visuo-spatial representations of physical situations that give us the ability to make judgments of the possible future outcomes of physical phenomena. Simulations are conducted by first sampling a noisy distribution containing values of perceptual and physical variables of the entities involved (J. R. Kubricht et al., 2017). The physical condition of entities are simulated at later times using sampled physical and perceptual variables of the environment, initially passed through a video game physics engine that approximates Newtonian mechanics. Predictions from each simulation are then retrieved to form a complete predicted judgment for that situation. Examples include: the direction of a falling tower of blocks and the relative angle of these falling blocks. The retrieved judgments are then compiled as a prediction distribution which serves to mirror human physical intuition of certain situations.

The use of physics engine in probabilistic simulation models of human intuitive physics contains uncertainties that are to be noted. Physics engines simulate approximate principles of Newtonian mechanics using noisy samples of perceptual and physical variables, thus deviations from normative principles occurs only due to the processing of these samples and not a result of a systematic bias in them. Furthermore, being simulation based, no analytic solution is obtained as scenes are generated using the previously generated ones.

3.8 Modern AI and Intuitive Physics

Let's take a look at some successful research studies which have shown that human physical predictions can be modeled using Machine Learning. Below are described the models which have successfully achieved the replication of human intuitive physics using approximate probabilistic simulation which opens the path to new avenues in computational research of humans physical scene understanding. Section 4.6.1 examines the success of approximate probabilistic simulation models aimed at representing the cognitive processes allowing people to infer on the physical outcomes and constraints of events involving rigid bodies. Section 4.6.2 explores the other experiments in which the aforementioned approach is included to model people's inferences about other objects and substances. Comparisons between these models and non-simulation models as well as models consisting of a combination of probabilistic simulation and data-driven aproaches will be explored.

3.8.1 Current Probabilistic Models of Intuitive Physics

In order to obtain a more complete understanding of the cognitive processes involved in human intuitive physics, the physical domain relating to rigid bodies was explored prior to other applications of inferential dynamics.

Earliest success was seen in (P. W. Battaglia, Hamrick, & Tenenbaum, 2013) which modelled people's predictive judgments of stacked blocks under various conditions and constraints by asking participants specific queries depending on the imposed constraints. Their model uses an "Intuitive physics engine" and simulations of approximated Newtonian mechanics, similar to what is discussed in 3.5 to model the responses of participants judgments of stacked block tower stability. Inputs to the "Intuitive physics engine" are sampled from a statistical distribution as to simulate a 3-D scene according to the objects present in it. The values from the distribution dictate the structural organization of the situation, the acting forces, and the general object properties that are initially constraint according to normative physics principles. They are subsequently modified by a set of parameters which account for mental imagery uncertainty, possible effects of unobserved forces, and latent properties of the physical structures in the situation. The modified inputs are then processed by a physics engine to simulate the dynamics in the physical situation to achieve a final scene output of approximated dynamics. These outputs are then collected as a set of predicates

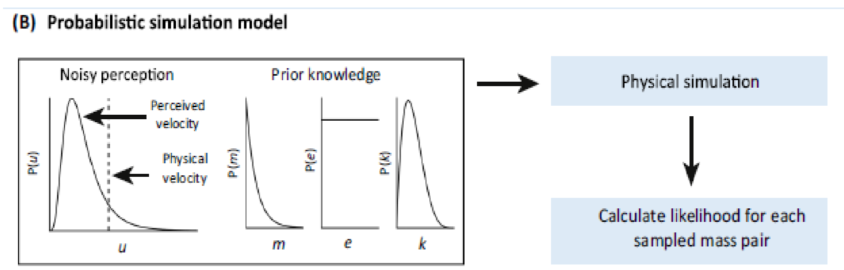

Figure 1: Probabilistic Simulation approach for the collision problem: The value of perceived velocity is sampled from a noisy distribution that is biased towards lower velocities, emulating the bias observed in people's judgments. The mass, relative velocity coefficient, and kinetic energy have values subsequently sampled from respective distribution. A video game physics engine then approximates Newtonian mechanics using these sampled values to determine the likelihood of certain mass ratios by comparing simulated mass ratios to those observed.

and applied to new simulation runs. Promising results have been obtained as human-level performance was achieved supporting that the underlying computational mechanisms present in human physical scene understanding is analogous to inference based on probabilistic simulation.

3.8.2 Inference of Non-Rigid Bodies Using a Physics Engine

The success seen in research geared towards modelling human physical scene reasoning lead the way for studies that extend the broadness of field applications of such models by applying it to new and more complex situations that we experience in our everyday lives.

We are exposed to the splashing, trickling, spreading, soaking, and pouring behaviours of liquids on a daily basis, and yet until now, no empirical evidence has shown the quantitative computational mechanisms which allow us to make predictions about these complex dynamic behaviours. One would be regarded as non-sensical if they believe that the underlying mechanisms that allow us to make such predictions are those of mentally computing fluid behaviour by solving Navier-Stokes partial differential equations. Therefore, it seemed evident to apply the previously used approach of modelling human judgment using a physics engine and probabilistic simulations as to determine if the way we reason with scenes of fluid dynamics is similar to our process of reasoning with scenes of rigid bodies (i.e block towers, colliding balls).

Bates (n.d.) shows that a model using probabilistic simulation of scenes involving liquid dynamics can be used to explain how humans reason with fluids on a daily basis. Through the use of a particle based intuitive fluid engine, their model assumes that people reason with dynamical scenes of liquids using particle based simulations permitting the approximate predictions of the possible later state of a system. The intuitive fluid engine consists of a particle based smoothed particle hydrodynamics computational method of simulating fluid dynamics combined with physical uncertainty of fluid particle initial positions. Given the task environment of predicting which side of obstacles will more fluid particles fall, participants judgments were compared with those made by two simulation models varying in fluid stickiness parameter. People's judgments were then compared with those made by the "intuitive fluid engine" simulations which provides supporting evidence for similar particle based simulations being instantiated in the brain when a human is presented with a scene involving fluids.

Research by Bates (n.d.) was extended to the task of container pouring angle by Kubricht et al. (2016) and improved in a multitude of realms. In this study, the task environment consists of participants making predictions on the pouring angle of containers varying in fluid volume and viscosity. The Fluid Implicit Particle/Affine Particle in Cell permits the simulation of physics using more than particle based simulations. The ommitance of the Smoothed Particle Hydrodynamics allows it to not have any stochastic processes accounting for perceptual noise. Therefore, simulations are conducted by passing noisy input variables of volume and viscosity to the FIP/APIC graphics simulator. Hence, this study shows that people consider the latent properties of fluids such as volume and viscosity when making predictions in dynamic scenes involving fluids. It also supports that the quantitative computational means in which people reason with physical scenes containing entities that exceed the complexity of rigid bodies is accorded by approximate mental simulation.

3.8.3 Data Driven Models

The success of probabilistic simulation as a model for human-level physical prediction has not yet reached the accord of all researchers. Some still remain firm in their conviction that the most efficient and accurate models will be of the traditional data driven type. In this section, the approaches to object interactions and motion prediction are explored in two studies. The execution of each experiment, their results, and the motives for developing such models despite prior success will be described to insure that not only a single account is delivered in this paper.

A study by Finn et al. (2016) is among those which attempts to reinforce the belief that motion prediction will be achieved using purely data driven models. This approach makes use of videos as training data for a model of motion prediction which allows a robot to predict pixel motion in a given video scene. It predicts the motion of pixels in subsequent frames by predicting a distribution over the pixel motion of the previous frames. Therefore, the motion of pixels in a previous frame is used to output the motion of pixels in the subsequent frame. This allows it to learn the dynamics of object interactions without the need for labelled data. Thus allowing it to gather virtually unlimited experience, without the need to store appearance information in the internal state of the model as it is directly available in the pixels of the previous frame. This model improves upon past models as it has a longer prediction range and uses raw video data.

Research by Levin et al. (2016) has lead to results serving as evidence towards the possibility that a data driven model of object interactions and physical prediction is more efficient than the probabilistic counterpart. Their approach aims to replicate the physical intuition that allows humans to plan future actions by predicting the outcome of physical events. The physical interactions of a robot was modelled using two convolutional neural nets trained on images of an initial state, an action on the objects in the initial state, and a final image state. These two neural nets work together to allow the robot to define certain actions with a corresponding change in visual state. What would have been a challenging supervised task is successfully made unsupervised task as the learning system consists of one convolutional neural network modelling forward dynamics(models the final image given an action and initial state image) and another modelling inverse dynamics(models the action given the initial and final states). The inverse model serves as the supervision that would be required to produce a model which consists of only forward dynamics modelling. These models work together to allow the robot to have a task oriented feature representation of its interactions with objects. Therefore, the robot is able to learn how objects interact through experience analogous to how infants develop physical expectations by playing with random objects. Levin et al. argues that this approach is of greater accuracy and efficiency than the simulation counterparts due to the errors that occur in the approximation of parameters that aren't present in deep learning models. Questions still remain about applicability in the real world and how training data could be more efficiently collected.

The main issue with data driven models lies in the enormous amounts of data required to train a model to perform decently before allowing it to proceed unsupervised. This equally shared concern has lead some to believe that the combination of data driven and probabilistic approaches may lead to the answer pertaining to the most effective model of motion prediction. And that is what is discussed in the next section.

3.8.4 Combined Physics Engine and Deep Learning Models

In this section, the remarkable success of modelling intuitive physics combining elements of simulation, probability theory, and deep learning to most efficiently replicate the cognitive mechanisms that underlie our physical judgments will be explored. Here is described a conceptual understanding of such models.

Yildirin et al. (2015) are among the first to have achieved successful results in replicating human level performance of physical judgments using a combined approach. Their model, "Galileo", operates under the assumption that humans possess a realistic 3-D physics which they use as part of a generative model to form complex internal representations of a physical scene, enabling intuitive predictions. The model is composed of a physics engine with a central generative model that yields inferences about the latent properties of objects in a physical scene given a visual input. These inputs consist of relatively simple yet rich physical scenes. The latent variables of volume, shape, friction, mass and position are estimated based on these inputs. Using these estimated properties, a realistic physics engine takes an initial scene configuration of the objects in the scene having the estimated properties and simulates the objects forwards in time. This forward simulation is the hypothesis for the generative model. Inference of latent physical properties in the hypotheses of the generative model takes place via Markov Chain Monte Carlo sampling of a distribution of values of these latent object properties. This probabilistic inference method effectively inverts the generative process by taking the output of each hypothesis and solving for the correct latent object properties that fit those of the physical scene input. Furthermore, the estimates of physical variables obtained from the physics engine are aggregated to serve as training data for a deep learning recognition model similar to Helmoltz machines (Dayan, Hinton, Neal, & Zemel, 1995). This recognition model then enables the generative model to make inferences of the physical evolution of scenes depicted in static images. Comparing this model with human performance has provided further evidence towards the probabilistic simulation account of human physical scene understanding.This is attributed to it accounting for the experiential elements involved as it not only closely matches human predictions but also human predictive errors.

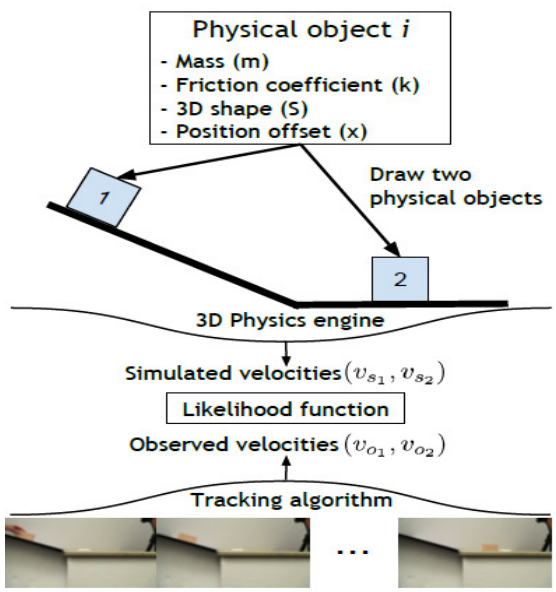

Figure 2: A hypothesis space consisting of the physical object representations is formed by the model. Given an input video, two objects along with their physical properties are drawn from this hypothesis space to simulate scene dynamics. Inference of physical properties taking place via Markov chain Monte Carlo simulation. Input and output scene dynamics are evaluated using their respective velocity vectors.

Battaglia et al.'s (2016) model reaches human level performance on numerous tasks that supersede the complexity of other's described earlier in this paper. In this model, as in others described, makes use of a physics engine, however, puts into question what a physics engine truly entails. The idea of what is typically thought of as a physics simulator takes on a new form in which a traditional physics engine, such as Bullet, is absent from it. In fact, it makes use of no simulation techniques of this type yet achieves the same task of propagating forwards in time the dynamics of a physical scene. This model, named, *interaction* network, accomplishes the task of a physics engine in which future object states are predicted through recursive application of physics laws, whilst also inferring the relations these objects have among each other. It does this using two neural networks: one of which is object-centric and the other relation-centric. The model thereby learns to distinguish objects and their relations in the implementation of neural networks, each having different tasks. It takes as input graphical representations of the state of a complex physical system, accounting for object physical properties such as mass, velocity and position. As it consists of two neural nets, the more abstract relations between these objects, such as gravity, rigidity, and spring constant are also taken as input. Given the initial state of a physical system consisting of objects and their relations, the interaction network infers its state at a future time step. It then uses this prediction as input for physical state prediction at the subsequent step, generating a scene output as such. This model has been shown to perform remarkably in complex tasks such as n-body problems, non-rigid body dynamics, and rigid body dynamics. Vast amounts of data are required to train this model, however, its success in 2-D problems provides evidence that additional training data could, in theory, allow for its application to more complex 3-D problems.

3.9 Remaining Unanswered Questions

Taking a step back, what can be concluded from all this research? The brain still remains the most unknown information processor studied by man. Yet answers to the most puzzling questions are believed to lie in modelling the capacities it grants in machines and learning systems. The enhancement of physical scene understanding in Machine Learning systems mark a step towards what could become a more comprehensive account of the neural circuitry in the brain. As it has been described previously, various studies have been found to support the probabilistic simulation account of human intuitive physics. However, the computational resources required to run such simulations remain unclear in their implementation in the brain. The probabilistic account also fails to represent how the human cognitive system can learn the physical properties of objects and their dynamics through experience. In sum, the enigma of the human brain persists, however, past conflicts between human intuitive physics and Newtonian mechanics have been effectively reconciled.

4 Artificial Intelligence

This section discusses attempts of reproducing human intelligence, introduced in Section 2, in digital space. Section 4.1 explores the history of artificial intelligence: from rudimentary forms to the modern alternatives. Section 4.2 presents the utility and basic functioning of Machine Learning, discussed further in Section 5.

4.1 History of Artificial Intelligence

What is Artificial Intelligence? Where does it come from? This section explores the root of Artificial Intelligence. In addition, this section examines how advancements in the field of technology allows for today's revolutionary Machine Learning.

Artificial Intelligence is not a modern concept. Philosophers and mathematicians have pondered if and how human cognition could be reproduced using basic logic for centuries. With the creation of mathematical disciplines such Boolean logic and Peano arithmetic, hypotheses suggested that human thought could be modelled using mathematics.

With the emergence of computational machines, Artificial Intelligence became reality. Research from mathematicians like Turing permitted human logic to be mimicked through circuitry. Inventions such as the Stochastic Neural Analog Reinforcement Calculator and Elektro, out of many others, pioneered what would become the Computer Sciences. As automation fueled industrialization, the idea of machines taking over humanity became, and still is, a societal fear.

The creation of simpler and more efficient computation systems allowed researchers and inventors to use machines for technological advancements and commercialization. Personal computing and arcades led to an Artificial Intelligence revolution. Video games such as Pac-Man and Mario Brothers created environments that could react to the player using complex rule-based Artificial Intelligences. These rule-based models are intelligent, seeing use in a variety of fields beyond gaming. This said, advancements during the last decade suggests that there could be another revolution in Artificial Intelligence.

Machine Learning is a field of Artificial Intelligence that relies on machine optimization to "learn" from data. Hypotheses were made over the years, as early as the 1960s, of the potential of a brain inspired structured model called Neural Networks. Due to limitations in computational power, today's Machine Learning algorithms only became relevant recently. This is attributed to improvements in hardware, mainly in the GPU, and optimizations in Neural Network algorithms, through libraries like TensorFlow and Pytorch. Today, Machine Learning, using Neural Networks, has become the norm for data processing and management.

4.2 Fundamentals of Machine Learning

What does it mean for a system to learn? Where and how can trained Neural Network be applied? Why do data scientists use Machine Learning extensively? Inspired by the introduction of Aurélien Géron's book, this section will focus on understanding Machine Learning and demonstrating why Neural Networks are revolutionizing data science.

4.2.1 Understanding Machine Learning

Géron describes Machine Learning as follows:

> Machine Learning is the science (and art) of programming computers so they can learn from data. (Géron, 2017)

In reality, Machine Learning is better described as an optimization problem. How can the input be transformed into the output with the least loss? Over the same problem, Machine Learning requires much more information than in human learning. This said, data alone doesn't allow Machine Learning;

> If you just download a copy of Wikipedia, your computer has a lot more data, but it is not suddenly better at any task. Thus, it is not Machine Learning. (Géron, 2017)

This considered, Machine Learning remains the best method of processing large amounts of data. Most problems originally solved using conditional logic can now be trained on Neural Networks. Moreover, Machine Learning can discover features that humans would struggle finding. Even if human models could be more accurate, Machine Learning models tend to take a fraction of the effort and time to train. Consider the following scenario:

> Imagine being given a database of a variety of living beings. This database consists of a large quantitative and qualitative feature list of species. Due to time constraints, previous teams of biologists were only capable of classifying the kingdom of a quarter of the species. In addition, a major bug led to a loss of a quarter of the features in database. The task consists of determining which species are mammals using the corrupted database.

Traditionally, a team of scientists would have to determine most mammalian features. This said, mammals consists of a large range of species from big and small to terrestrial and aquatic. Let it be assumed that mammals are only tetrapods. Most terrestrial mammals would be correctly classified, but nearly no aquatic mammals would be identified. In addition, there is the risk of classifying amphibians as mammals. And what happens when species are missing the "leg number" feature in the database? You hope that your variety of predictions will fill in the blank. Thus, traditional methods of classification tends to be lengthy and inefficient.

Today, a team of scientists would analyze and cleanup the data. From the data structure, a Machine Learning architecture would be determined. The learning method, in this case supervised training, is then selected. Through rigorous hyperparameter selection, a model can be trained to recognize mammals. This implies that Machine Learning determines mammalian features without human interventions. This often yields models with a better understanding of large databases than our intuitive models. This allows trained models to classify odd mammals like whales, which our previous "tetrapod" assumption wouldn't. Thus, Machine Learning becomes a tool that assists scientists with large amounts of data that would require massive teams to extrapolate correlations between features.

To sum up, Machine Learning is a technique that allows computers to understand data. Depending on the size of databases, various architectures, usually Neural Networks, can be applied to yield desired outputs, from labels to images. In most cases, Machine Learning "black box" model is more robust and efficient than a human based model. This is why we can observe a shift towards Machine Learning in the scientific community.

4.2.2 Types of Machine Learning

Training great Machine Learning models requires thoughtful architecture structuring. This decision is based on the data given to the system and the data expected by the system. Thus, understanding the problem is essential to select proper models.

Initially, the learning method should be determined. This falls in four major categories. Supervised learning trains on labeled data to make predictions. A basic example is classifying cats and non-cats. Unsupervised learning trains on unlabeled data to make predictions. Online shopping websites use this to group clients with similar interests. Semisupervised learning trains on both labeled and unlabeled data to make predictions. Photo galleries use this to link photos containing the same person, later prompting the user to name that person. Lastly, Reinforcement learning trains agents to perceive and act upon their environment. Video Game "bots" use this strategy to solve levels without prior human knowledge. An example of this can be found in Mari/o (SethBling, n.d.).

Next, the training method should be determined. This is falls in two categories. Batch learning, or offline learning, uses all the data to train. This leads to a stable but unchanging model. This means that new models need to be trained from scratch on old and new data. Online learning uses segments of the data. This allows for a model with a constant flow of data over many processors, generating a rapid and adaptive network. The main disadvantage of online learning is that it is highly susceptible to bad data.

Last, the generalization method should be determined. Instance-based learning labels new data according to the nearest learnt point. Multi-object classifiers uses this method to match new objects with learnt objects. Model-based learning uses "functions" to make predictions. Linear models apply this technique to predict outputs of linear data.

In brief, Machine Learning requires a careful examination of data. Understanding the problem at hand allows for the best selection of techniques to train an efficient model. However, this selection process tends to be lengthy and has the risk of complications during training.

4.2.3 Complications with Machine Learning

Model selection is a major step in Machine Learning. Regardless of the effort, a model can still fail during training. This illustrates the often overseen complexities of Machine Learning; models often need extensive tweaking to avoid the reoccurring difficulties.

First, Machine Learning is sensitive on data. Any irrelevant or bad data leads to difficulties in training. This means an added training time or a model that cannot converge. In addition, models often need large databases to generalize. Lacking information often leads to a model making bad predictions due to a bad understanding.

Second, Machine Learning needs extensive tweaking to represent the data correctly. Overfitting can occur when a model is given too much variability. This leads to models that perfectly represents old data but unable to generalize to new data. Underfitting is the opposite; the model is given too little variation. This leads to most data being misrepresented.

To summarize, selecting a model that fits your data is not the only complexity of Machine Learning. Understanding how to deal with complications in training is essential for any Machine learning model to converge. One that can master this section, in addition to the others preceding, could be considered a Machine Learning expert.

5 Deep-Physics Model

What is Deep-Physics? This section focuses on understanding our Deep-Physics model. Section 5.1 is a formal proposal, in English and mathematical form, on how to solve intuitive physics using Artificial Intelligence. Sections 5.2, 5.3, and 5.4 discusses the possible Neural Network models able to solve the algorithm discussed in Section 5.1.2. Section 5.5 theorizes which physics problems a Deep-Physics model should be able to solve. Section 5.6 proposes ways to visualize trained Deep-Physics models in real-time.

5.1 Proposal

How do humans perceive physics? Section 3 presents the human "intuitive" ability of understanding their environment. How can Artificial Intelligence replicate this innate human ability? One could create a vast conditional model, the basis of typical physics engines, but is this truly how we render physics? Deep-Physics attempts to more closely replicate human's perception of physics by using Machine Learning.

Deep-Physics is a collection of Neural Networks with the goal of replicating physics. In principle, Deep-Physics can be summarized as: feed objects and environments and return the physics. Problems arise when considering the variability of the models' parameters. What data and type of Neural Network should be used? There isn't one perfect solution to this question. We decided to base Deep-Physics on the algorithm mentioned in Section 5.1.2. This limits our selection to numerical data and a handful of Neural Networks.

5.1.1 Data Structure

There are various data types that could be used to train a physics Neural Network. We've concluded that using numerical and simulated physics would be the most efficient solution for Deep-Physics.

Theoretically, visual data could be used for training, but this adds a convolution layer and potentially increases the training time. In addition, there already exists outstanding visual networks that can visualize environments, but aren't able to render the physics behind these environments. Thus, we propose that chaining these visual networks with Deep-Physics, as in 3, is the most efficient solution.

We also have to consider that there exists massive training sets for both visual and numerical physics. This said, creating virtual environments with physics engines allows to generate large amounts of objects with drastically different parameters. Even if physics engines will never match actual data, the approximations are good enough to fool human physics understanding. Therefore, these simulations would be in essence the same as real-world data.

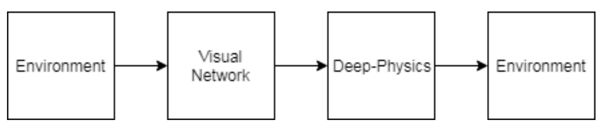

Figure 3: Diagram demonstrating the hypothetical processing of physics. An environment is deconstructed to basic objects by a visual network. Deep-Physics applies the physics and reintroduces the objects to an environment.

5.1.2 Algorithm

Let O be an object with position $\vec{P}$, velocity $\vec{V}$, acceleration $\vec{A}$, and shape class C. Let environment E be a collection of O. Let network N be Deep-Physics, a neural network that replicates physics over a collection of short periods of time dt. Let ^{P} be an element generated by a physics engine. Let ^{N} be an element generated by N. Let O_{i+n} be an object that had a collection of small variation $d\vec{P}$ and $d\vec{V}$ applied over dt.

Pick E from a training set. A random object O_i^P is taken from E. O_i^P is fed separately from E to N. N applies changes to O_i^P such that it becomes O_{i+n}^N

according to E. $d\vec{P}$ and $d\vec{V}$ of O^N_{i+n} and O^P_{i+n} are compared using an error function.

5.2 Deep-Physics and DNNs

A Dense Neural Network (DNN) could be a simple yet efficient model to make predictions of physics of a singular object over a small time-step. Feeding back predicted objects to the trained model could allow for a recurrent structure that doesn't depend on time or multiple previous states. This makes computations quick, as they are simple feed-forward models, and allows multiple networks to run in parallel, yielding a physics engine behaviour.

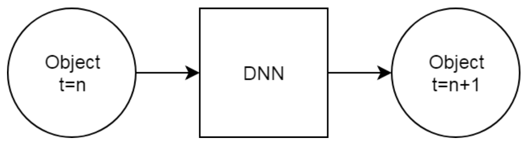

Figure 4: Graph demonstrating the structure of a Deep-Physics DNN. An object with it's various states is passes through the network yielding the next state. This allows the object to be passed recurrently through the network to return a path.

5.2.1 Model Structure

The DNN variance of Deep-Physics can be split in three layers. The input layer is an object array with the environment array concatenated to it. The dense hidden layer is composed of multiple ReLU activated layers of funnelling size to allow for an efficient optimization. The output layer is another dense layer of the object array size using a tanh activation, allowing for negative results. The model is compiled with an Adam optimizer and MSE loss function.

5.3 Deep-Physics and GANs

First introduced by (Goodfellow et al., 2014), the Generative Adversarial Network (GAN) has become a staple of complex data generation. The model's efficiency is often explained using these examples:

> The job of a money counterfeiters is to make the best fakes they can. The job of a cop, on the other hand, have to improve their detection techniques. This forces both parties to compete and learn faster. In another perspective, a biologist would call this network competition an "Evolutionary Arms Race".

As many physics models can be described with few parameters, GANs should be able to generate "fake physics". This said, the complexity or simplicity, depending on the physics model, could lead to different structures of the network to learn much faster than the others. Why and how this could be an issue will be explored further in the following sections.

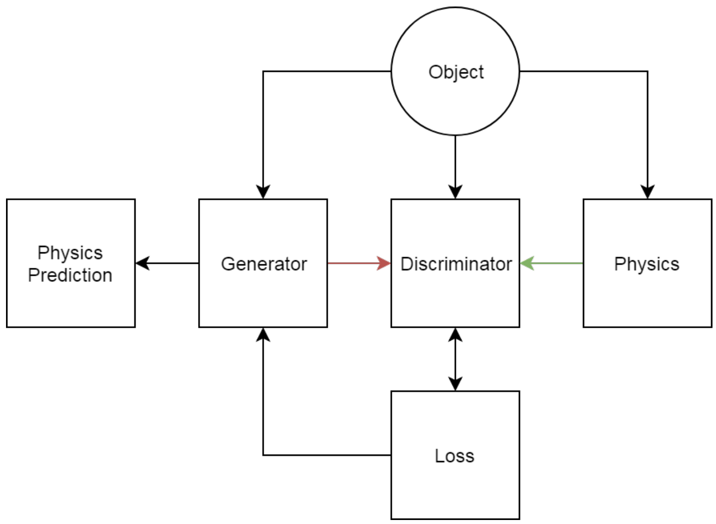

Figure 5: Graph demonstrating the structure of a Deep-Physics GAN. Objects are fed to the Generator and compared to Physics models using the Discriminator. The loss of the Generator/Discriminator (Adversarial Network) is determined by how similar the generated data is to the actual data. The loss of the Discriminator Network is determined on the accuracy of classification between generated and actual data. After training, the Generator should yield data very similar to actual physics, demonstrating learning of physics intuitively.

5.3.1 Adversarial Network

The Adversarial Network is the generative model of a GAN. In typical scenarios, noise space is fed to this network to generate images. We propose that we could use previous physics states to make future previous states. This allows for the Adversarial Network to have the same structure as the DNN described in Section 5.2.

The difference between the Deep-Physics GAN and DNN is in the end layers. The goal of the Adversarial Network is now to get labeled as a “false-positive” by the Discriminator Network, Section 5.3.2. As the complexity of physics models varies, there is a possibility that simple physic models can be identified really easily by the Discriminator, preventing the Adversarial Network from learning. Given this constraint, a generator that passes the training phase should yield physics more accurate than a simple DNN.

5.3.2 Discriminator Network

The Discriminator Network is the labeling model of a GAN. Given real and generated information, in this case physics, the Discriminator Network should be capable of distinguishing between them. For Deep-Physics, an effective Discriminator Network should understand the fundamentals of the physics models used for training. This should be achievable using a DNN structure that flows into a single output node, yielding a loss to the Adversarial/Discriminator Network.

Note, even if this research focuses on producing an effective Generator Network, training an effective Discriminator Network could potentially lead to a better understanding of intuitive physics. Testing samples, like video games, and getting the Discriminator's "physics rating" could potentially allow us to understand the limits of physics simplifications through physics simulations.

5.4 Deep-Physics and RNNs

Recurrent Neural Networks were first introduced in the 1980. Today, most RNNs, including those part of the TensorFlow library (Abadi et al., 2015), use the LSTM variant, introduced by (Hochreiter & Schmidhuber, 1997). These networks are effective at finding patterns in sequential data; models that rely on previous states to make predictions on future states. This makes RNNs effective at language processing, speech recognition, and sequential data generation. Given Deep-Physics' goal of physics path prediction, the RNN structure should be capable of predicting T number of steps given initial conditions of environments.

5.4.1 Model Structure

The RNN variance of Deep-Physics can be split in four layers. The input layer is an array of objects N in their initial states. The LSTM hidden layer is composed of LSTM cell at every time-step. This allows to feed environment states (forces, collisions, etc) throughout the "simulation". The path generated by the LSTM hidden layer is fed to a dense layer, returning normalized data. The normalized is returned by the output layer where the path for T steps is given for each N object.

5.5 Applications

Understanding Neural Network models and data structures is one thing, but what makes an effective Deep-Physics model? There are numerous problems that could be chosen, but which ones should universal physics model be able to solve? This section compiles a short list of problems that were deemed necessary for an effective Deep-Physics model to solve. These were chosen in consideration of what makes physics engines effective at fooling humans' intuitive physics: motion, forces, and collisions.

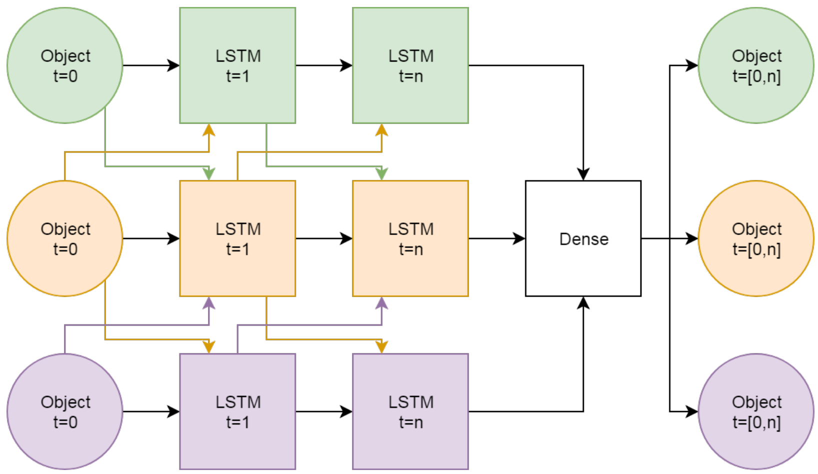

Figure 6: Graph demonstrating the structure of a Deep-Physics RNN. Every object in an environment can be cross fed the previous LSTM of every other object to the current object LSTM. This allows for a path prediction that considers the state of other objects. In addition, the environment can be fed to every time step to have changing forces.

5.5.1 Kinematics

Kinematics is the simplest branch of dynamic systems and Newtonian Physics. This includes a list of simple equations found in basically every physics textbook. A Deep-Physics model is expected to be able to solve these equations, or an equivalent of these equations, efficiently to be considered a "dynamic system solver". This is our take on the problem:

A body B in $\mathbb{R}^n$ is defined by a position vector $x(t)$, a velocity vector $\dot{x}(t)$, and an acceleration vector $\ddot{x}(t)$ with initial states x_0 and $\dot{x}_0$. Body B's motion should follow these equations:

$$\dot{x} = \int \ddot{x} + \dot{x}_0 \tag{6}$$

$$x = \iint \ddot{x} + \dot{x}_0 t + x_0 \tag{7}$$

Where the velocity and acceleration are differentiable over t_0 to t_n. These equations can also be expressed as system of coupled first order differential equations for numerical approximation:

$$\frac{d\dot{x}}{dt} = \ddot{x} \tag{8}$$

$$\frac{dx}{dt} = \dot{x} \tag{9}$$

Where the initial conditions are set according to x_0 and $\dot{x}_0$. Although these equations seem trivial, they should not be disregarded. They will serve as a basis for our Deep-Physics model to solve following problems.

5.5.2 N Body Problem

The N Body Problem is a complex application of Newton's Laws. The problem is an introduction to the Chaos Theory branch. A Deep-Physics model able to solve this problem demonstrates an understanding in passive interactions between objects. The problem can be modelled using coupled differential equations considering the number of the bodies in the system N:

Body B in $\mathbb{R}^{\mathrm{n}}$ can be defined by a position vector x, a velocity vector $\dot{x}$, and an acceleration vector $\ddot{x}$. The force between two body can be found using Newton's Law of Gravitation:

$$F_i = \frac{Gm_i m_j}{\vec{r}_{ij} \cdot \vec{r}_{ij}} \hat{r}_{ij} \tag{10}$$

Where F_i is the force on B_i by B_j in N, m_i and m_j are the masses of B_i and B_j in kg, $\vec{r}$ is the displacement vector from B_i to B_j in m, and $\hat{r}$ is the direction of $\vec{r}$. The net force acting on B_i is an extension of equation 10:

$$F_i = \frac{Gm_i m_1}{\vec{r}_{i1} \cdot \vec{r}_{i1}} \hat{r}_{i1} + ... + \frac{Gm_i m_N}{\vec{r}_{iN} \cdot \vec{r}_{iN}} \hat{r}_{iN} \tag{11}$$

Where F_i is the net force on B_i by $[B_1, ..., B_N]$ in the system. Using $F = m\ddot{x}$, the equation can be simplified to yield the acceleration of B_i:

$$\ddot{x}_i = \frac{Gm_1}{\vec{r}_{i1} \cdot \vec{r}_{i1}} \hat{r}_{i1} + ... + \frac{Gm_N}{\vec{r}_{iN} \cdot \vec{r}_{iN}} \hat{r}_{iN} \tag{12}$$

The position, velocity, and acceleration vector of B_i can be coupled to yield a list of first order linear differential equations:

$$\frac{dx}{dt} = \dot{x} \tag{13}$$

$$\frac{d\dot{x}}{dt} = \frac{Gm_1}{\vec{r}_{i1} \cdot \vec{r}_{i1}} \hat{r}_{i1} + ... + \frac{Gm_N}{\vec{r}_{iN} \cdot \vec{r}_{iN}} \hat{r}_{iN} \tag{14}$$

Where $\frac{dx_n}{dt}$ is the derivative of the position vector equal to the velocity vector, and $\frac{d\dot{x}_n}{dt}$ is the derivative of the velocity vector equal to equation 12. Although these first order linear differential equations can be solved analytically, a numerical solver like Runge-Kutta yields a close approximation of the lengthy system of equations.

5.5.3 Collisions

Collisions between bodies can be "intutively" solved by humans. It seems vital for a universal physics model to solve this problem. The complexity of a Machine Learning model to understand this problem lies in the bodies themselves. A Deep-Physics model able to solve this problem demonstrates an understanding of active interaction between objects, supporting an understanding of object properties like mass and geometry. This phenomena can be described using Newton's Conservation of Energy:

$$m_i\dot{x}_i^0 + m_j\dot{x}_j^0 = m_i\dot{x}_i^1 + m_j\dot{x}_j^1 \tag{15}$$

Where the collision is elastic and m_n and $\dot{x}_n^i$ are the masses and velocity of body B_n at time i.

$$m_i\dot{x}_i^0 + m_j\dot{x}_j^0 = (m_i + m_j)\dot{x}_{ij}^1 \tag{16}$$

Where the collision is inelastic and $\dot{x}_{nm}^i$ is the combined velocity of body B_n and B_m at time i.

5.6 Visualizations

A trained Deep-Physics model should substitute physics models. Although using Deep-Physics purely numerically should be possible, Deep-Physics as graphs or as real-time physics simulations demonstrates the future of "human-free" physics. Although any programming language should be able to use Deep-Physics models, this section compiles languages with "easy" access to Machine Learning libraries based on TensorFlow. Furthermore, this section suggest techniques to visualize data from these TensorFlow based libraries.

5.6.1 Python

Python has no doubt become the programming language of the scientific community. Scholars, engineers, and computer scientists (of many) use the ever expanding collection of libraries from work varying from research to day-to-day application. One field of study that is still remains almost purely on Python is Data Science.

Data Scientists, varying from data analysis to Machine Learning, require libraries like pandas, numpy, and TensorFlow on a day-to-day basis. To this day, these libraries still mostly reside on Python. Thus, training and using a Deep-Physics model is almost fully limited to Python. Luckily, Python offers a variety of libraries for visualizing data.

Matplotlib, a library based on Matlab, can create a variety of graphic information. Using this library, Deep-Physics can be both plotted as a path and an animation of motion. This library, in combination with numpy, also allows to visualize Deep-Physics: in three dimensions, as the difference with real physics, and many more. This said, Matplotlib doesn't support user interactions, requiring more game-centered libraries, like Pygame, for the physics engine property.

This said, TensorFlow/Keras models can also be converted to TensorFlow.js format to use with JavaScript.

5.6.2 Javascript

JavaScript, brother to Java, is a versatile language. Although often associated with web-design, JavaScript also contains a variety of useful libraries. Recently, TensorFlow was ported to JavaScript, called TensorFlow.js, allowing for web-based Machine Learning using WebGL. In combination with P5.js, Machine Learning can be modelled in real-time considering user input.

P5.js (McCarthy, n.d.), is a JavaScript version of the Java based library Processing. This library allows for fast and easy programming of interactive visuals. With a few lines of code, the physics from Deep-Physics can be modelled, allowing the user to change parameters in real-time. One could potentially create a game running on Deep-Physics, demonstrating the potential of a Deep-Physics physics engine.

6 Results

This section summarizes the results of some of the applied theory on Deep-Physics from Section 5. Results were separated according to the type of Neural Network used. For formatting purposes, graphs and other visual elements, all including captions, were placed in Section 8. All code pertaining this research can be found on `https://github.com/alexandre-lavoie/deep-physics`.

6.1 DNNs

The DNN variance of Deep-Physics was concluded to be an efficient and reliable solution for simple physic models. Using Keras, the DNN was trained on Newtonian kinematics, as discussed in Section 5.5.1, and yielded paths with physics-like properties. Training more chaotic physics systems, like the N Body Problem from Section 5.5.2, was inconclusive during this training period.

Training the Deep-Physics DNN on physics with basic properties required a deeper neural network than expected. A compact network of three layers, totalling around a thousand nodes, produced continuous paths but diverged from actual physics, even with a couple hours of training. Deeper networks of five layers, with around three times the amount of neurons the compact network contained, produced very promising data within an hour of training. Using this current model structure would require massive networks for chaotic systems, which testing seemed to validate. This limited the research of Deep-Physics DNNs to training with Newtonian Kinematics.

Using P5.js, the trained Deep-Physics DNN was applied in a variety of physics scenarios in real-time. Projectile motion and acceleration towards a point, all with random initial values, were tested and yielded physics-like visuals. With closer inspection, the paths produced by the Deep-Physics DNN had regions of higher inaccuracy, more noticeable around the origin and boundaries. This was hypothesized to be caused either by the training or the lack of training points in these regions.

Using Matplotlib, the trained Deep-Physics DNN was compared to actual physics over thirty time steps. Some paths had substantial error over the short period, which our intuitive physics could not detect. This suggest again that there exists small regions of inaccuracy, which during real-time simulation is hard to perceive.

6.2 GANs

The GAN variance of Deep-Physics yielded inconclusive data throughout the research period, leading to its abandon. This said, given the GAN was implemented using Keras, there could be issues on the application level. Taking this in consideration, this section will discuss the problems encountered and how it could potentially be solved using other techniques.

Figure 10 demonstrates the loss of both the Adversarial and Discriminator Network over epochs. Both network appear to compete, yet the Adversarial loss increases over time. This seems to suggest, as hypothesized in Section 5.3, that the simplicity of the network leads to the Discriminator learning much quicker than the Generator. Using basic hyperparameter tweaking, learning rate changing and hidden layer scaling did not solve this issue. This seems to suggest a deeper customization of the GAN.

Keras is a great library for high level Machine Learning. This said, results from the GAN suggests the need to use the lower level TensorFlow library. Seeing results from other Keras networks (DNN and RNN) led to a focus towards those networks. This research does not conclude that this network structure isn't viable, the GAN only seems to require customization, which wasn't possible during this research period. The "physics rating" Discriminator hypothesis from Section 5.3.2 still should be discussed.

6.3 RNNs

The RNN variance of Deep-Physics was concluded to be an effective model for simple physics with recurrent behaviour. In addition, RNNs were shown to be able to solve simple systems of differential equations. Using Keras, the RNN was able to produce lengthy paths, 100 points, using the technique described in Section 5.4. Larger systems of differential equations, as in chaotic motion, were inconclusive in the training period.

Paths successfully generated with the RNN had high accuracy and precision when compared to solver results. These paths were learnt in a brief training period of ten epochs using ten thousand generated paths. This mentioned, not all systems of equations, even the simpler ones, converged. This suggests that some systems have either more complex motion or require a more careful normalization.

Although entirely contained in Keras during training, RNN yielded results that would suggest this model being the most effective at solving the Deep-Physics proposal. The generated paths would follow the solver solution much closer than DNNs, and at occasions, appears to course correct. This is hypothesized to be due to the recurrent property of RNNs, which is also found in physics. This further demonstrates the utility of previous data to make predictions, leading this research to conclude RNNs to be more effective than DNNs and the most probable to fully solve the Deep-Physics proposal.

7 Conclusion

Deep-Physics suggests to be a potential alternative to physics models. Results from the Deep-Physics DNN and RNN demonstrates the possibility of using these Neural Network architectures as physics simulations, accurate enough to fool our intuitive physics. Lack of results in more complex scenarios, like in the N Body Problem and Collisions, prevents this research from concluding the application of Deep-Physics beyond Newtonian Kinematics and Second Order System of Differential Equations. Issues encountered in more complex physics scenarios suggests the need for more careful data generation and normalization, Neural Network hyperparameter tweaking, and model structure manipulations. With these issues resolved, an RNN appears to be the best method of generating complex physic paths. If this model could learn physics efficiently with a manageable computation load, Deep-Physics could become the future of rule-less physic engines.

8 Appendix

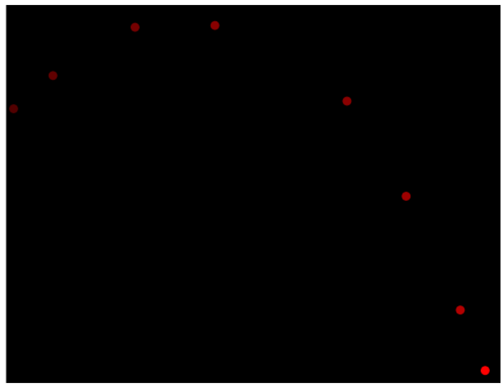

Figure 7: Image demonstrating projectile motion using P5js and DNN. Images of the particle were taken at random intervals and stitched together.

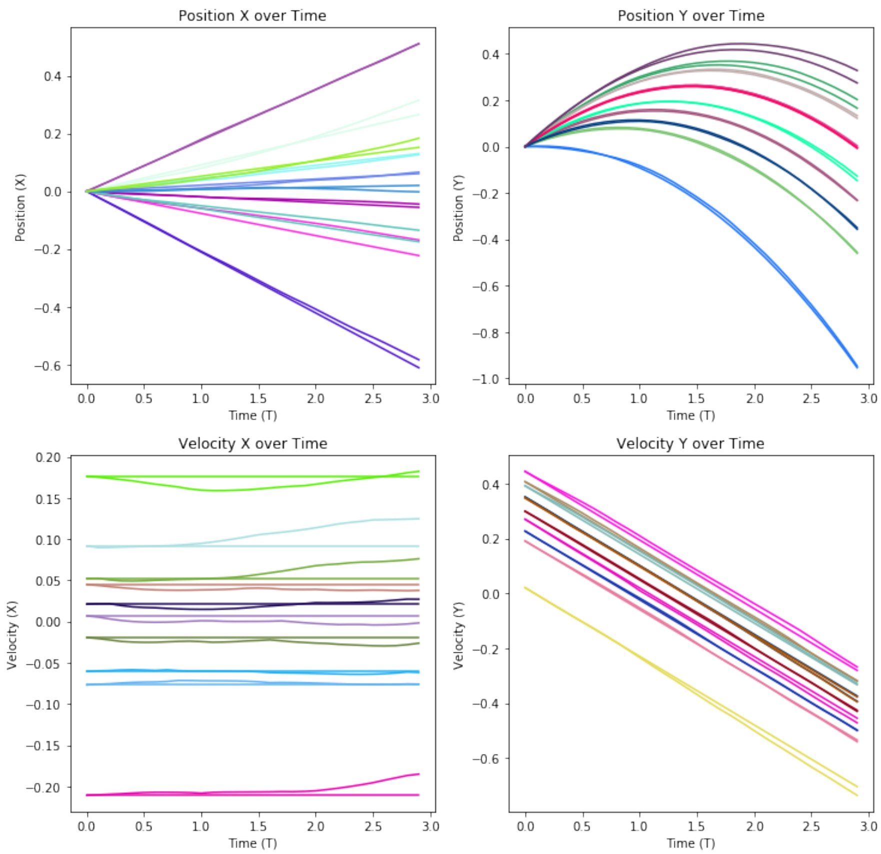

Figure 8: Graphs demonstrating the motion of the DNN Deep-Physics model after one hour of training. Each same colored line represents a randomly generated object with classical motion and model motion applied.

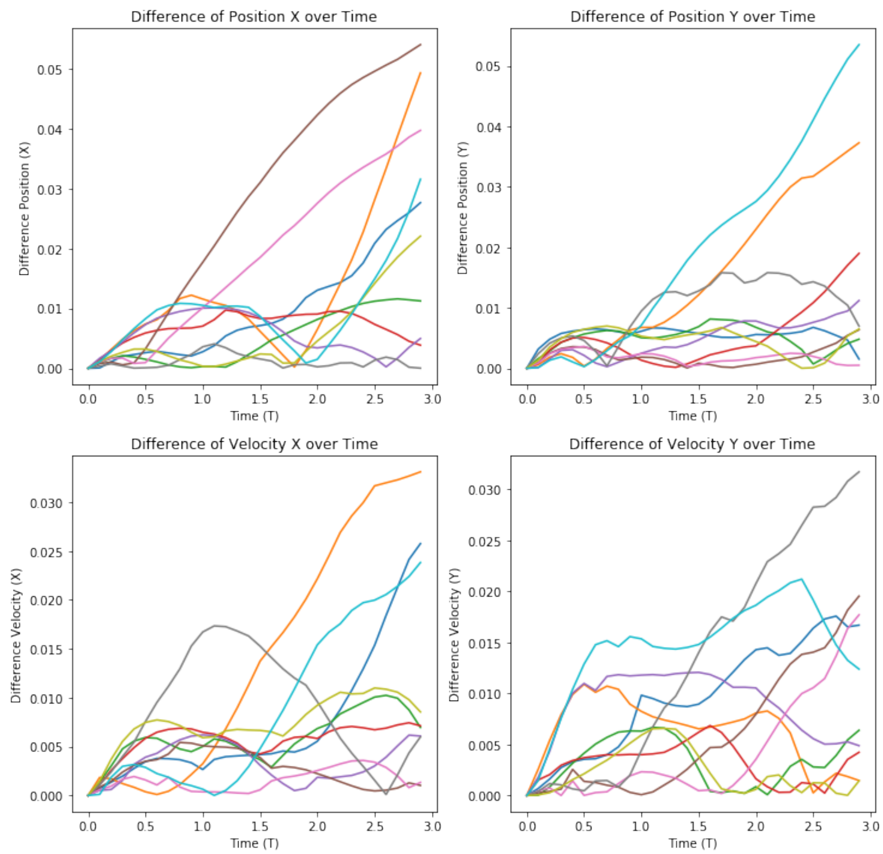

Figure 9: Graphs demonstrating the error of the DNN Deep-Physics model after one hour of training. Each line represents the difference in motion of an object a randomly generated object.

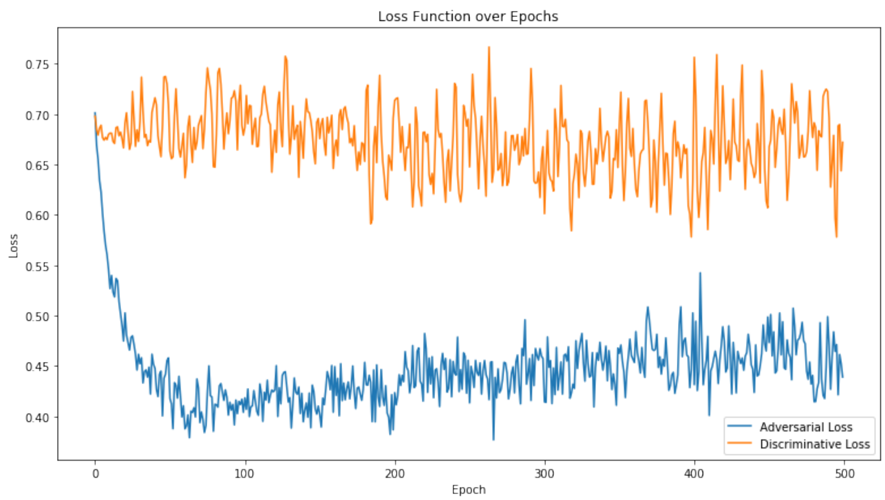

Figure 10: Graphs demonstrating the loss of both the Adversarial and Discriminative network. Competition is occurring between the models, yet the Generative network does not yield expected results.

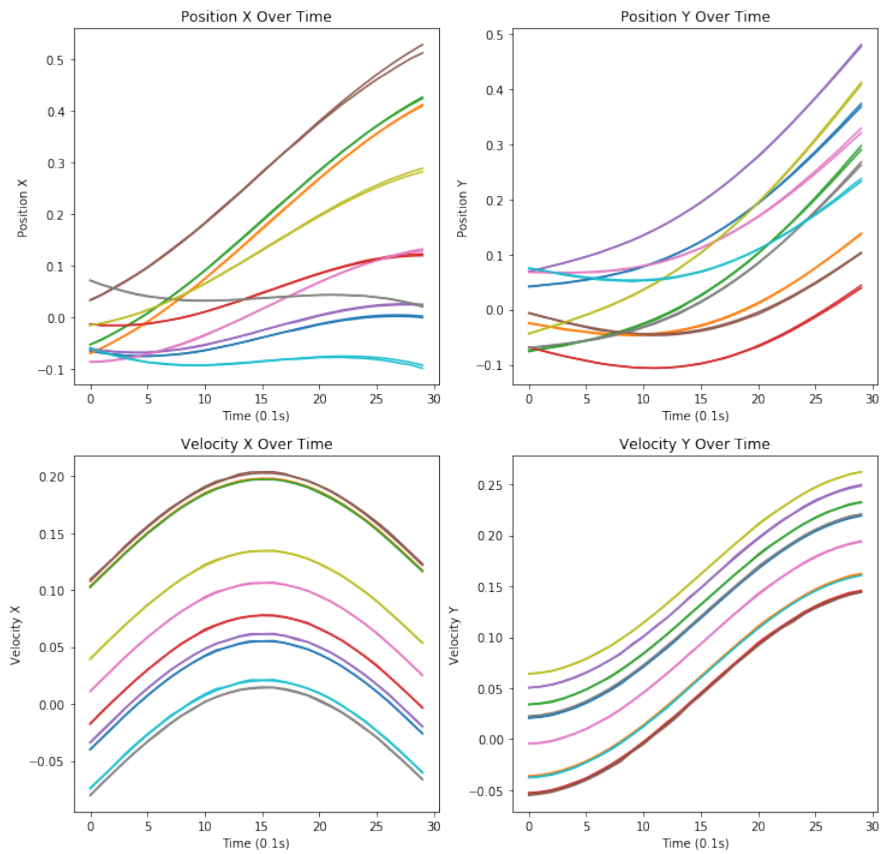

Figure 11: Graphs demonstrating harmonic acceleration motion of the RNN Deep-Physics model after a few minutes of training. Each same colored line represents a randomly generated object with classical motion and model motion applied.

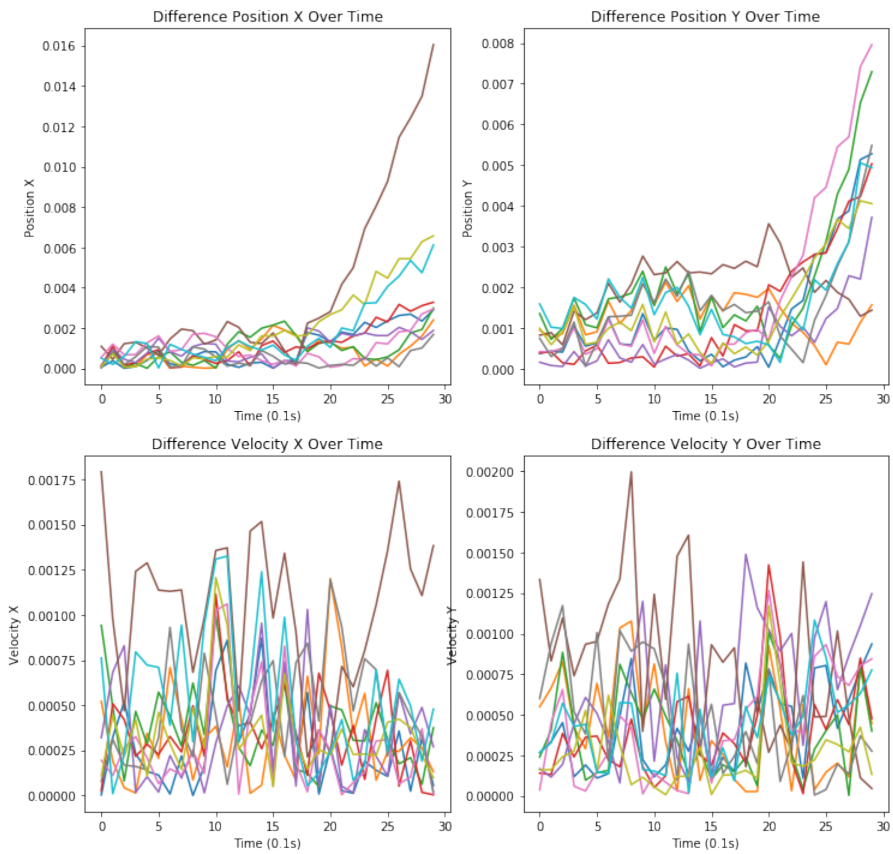

Figure 12: Graphs demonstrating the error of the classical motion RNN Deep-Physics model after a few minutes of training. Each line represents the difference in motion of an object a randomly generated object.

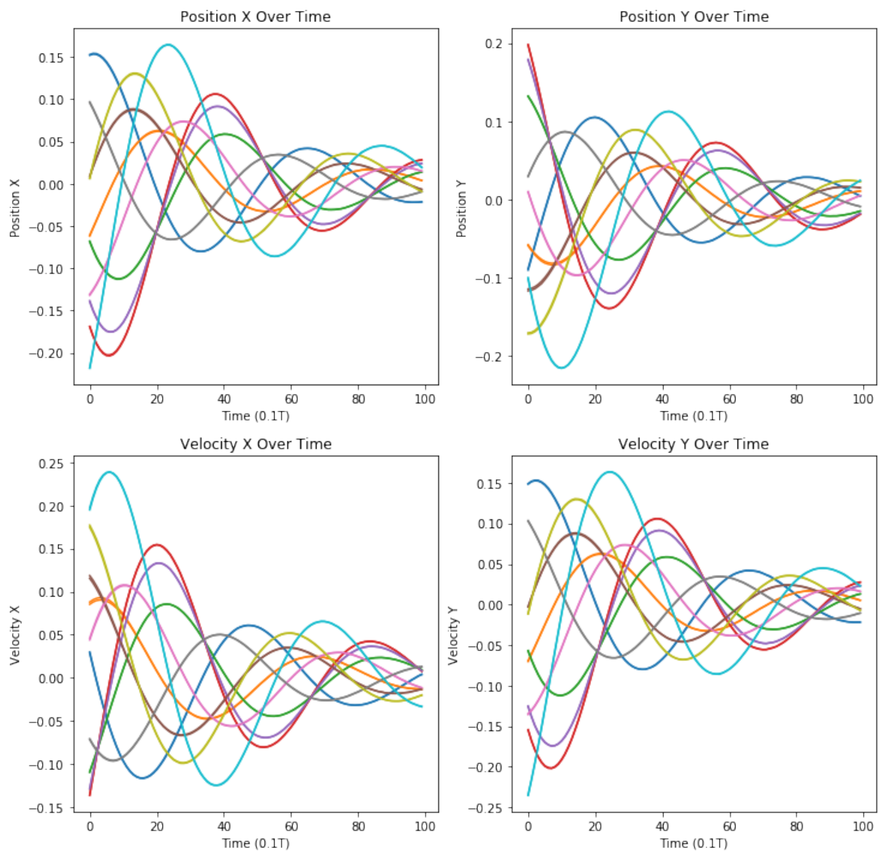

Figure 13: Graphs demonstrating motion of the system of differential equations: x'=-0.5x-y and y'=sin(x). Each same colored line represents an object with a random initial condition. In 10 epochs, around a minute of training, RNN Deep-Physics generated paths very close to the solver solution.

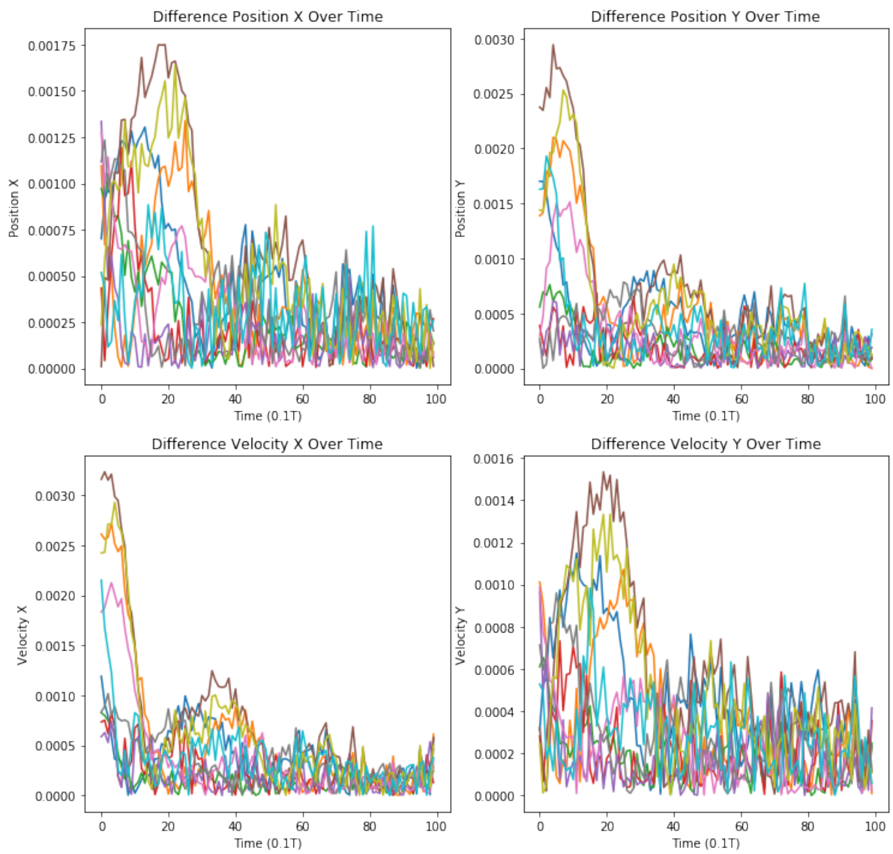

Figure 14: Graphs demonstrating the error of the system of differential equations RNN Deep-Physics model after a few minutes of training. Each line represents the difference in motion of an object a randomly generated object.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from `https://www.tensorflow.org/` (Software available from tensorflow.org)

Agrawal, P., Nair, A., Abbeel, P., Malik, J., & Levine, S. (2016). Learning to poke by poking: Experiential learning of intuitive physics. *CoRR*, *abs/1606.07419*. Retrieved from `http://arxiv.org/abs/1606.07419`

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition*, *23*(1), 21–41. doi: 10.1016/0010-0277(86)90052-1

Baillargeon, R. (1987). Object permanence in 3- and 4-month-old infants. *Developmental Psychology*, *23*(5), 655–664. doi: 10.1037/0012-1649.23.5.655

Barber, D. (2011). *Bayesian reasoning and machine learning.* Cambridge University Press.

Battaglia, P., Pascanu, R., Lai, M., Rezende, D. J., et al. (2016). Interaction networks for learning about objects, relations and physics. In *Advances in neural information processing systems* (pp. 4502–4510).

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

Breedin, N. J. C. S. D. (1994). Constructing naive theories of motion on the fly. *Memory Cognition*, *22*, 474–493. doi: 10.3758/bf03200871

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The helmholtz machine. *Neural computation*, *7*(5), 889–904.

Finn, C., Goodfellow, I. J., & Levine, S. (2016). Unsupervised learning for physical interaction through video prediction. *CoRR*, *abs/1605.07157*. Retrieved from `http://arxiv.org/abs/1605.07157`

Fischer, J., Mikhael, J. G., Tenenbaum, J. B., & Kanwisher, N. (2016). Functional neuroanatomy of intuitive physical inference. *Proceedings of the national academy of sciences*, *113*(34), E5072–E5081.

George, P. (1945). *How to solve it: a new aspect of mathematical method.*

Géron, A. (2017). *Hands-on machine learning with scikit-learn and tensorflow: concepts, tools, and techniques to build intelligent systems.* " O'Reilly Media, Inc.".

Gibson, J. J. (1966). The senses considered as perceptual systems.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).

Gottfredson, L. S. (1997). *Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography.* Citeseer.

Haier, R. J. (2017). *The neuroscience of intelligence.* Cambridge University Press.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in cognitive sciences*, *8*(6), 280–285.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780. Retrieved from `https://doi.org/10.1162/neco.1997.9.8.1735` doi: 10.1162/neco.1997.9.8.1735
Humans predict liquid dynamics using probabilistic simulation. (n.d.).
Kubricht, J., Jiang, C., Zhu, Y., Zhu, S.-C., Terzopoulos, D., & Lu, H. (2016). Probabilistic simulation predicts human performance on viscous fluid-pouring problem. In *Cogsci.*
Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in cognitive sciences*, *21*(10), 749–759.
Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.
McCarthy, L. (n.d.). *P5js.* Retrieved from `https://p5js.org/` (Software available from p5js.org)
MCCLOSKEY, A. G. B., M.; CARAMAZZA. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science*, *210*, 1139–1141. doi: 10.1126/science.210.4474.1139
McCloskey, M. (1983, 01). Naive theories of motion. *Mental Models*.
Myron L. Braunstein, E. C. C., & (Auth.), M. P. F. (1976). *Depth perception through motion* (illustrated edition ed.). Elsevier Inc, Academic Press Inc.
N. Eilan, R. M., & (eds.), B. B. (1993). *Spatial representation.* Oxford:Blackwell.
Runeson, S. (1977, 01). On visual perception of dynamic events.
Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological Review*, *120*(2), 1–77. doi: 10.1037/a0031912
SethBling. (n.d.). *Mari/o - machine learning for video games.* Retrieved from `https://www.youtube.com/watch?v=qv6UVOQ0F44`
Squibb, S. J. P. H. R. (1990). Algebra mal-rules and cognitive accounts of error. *Cognitive Science (Elsevier Science)*, *14*, 445–481. doi: 10.1016/0364-0213(90)90019-s
Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. Retrieved from `http://science.sciencemag.org/content/331/6022/1279` doi: 10.1126/science.1192788
Todd, J. T., & Warren Jr, W. H. (1982). Visual perception of relative mass in dynamic events. *Perception*, *11*(3), 325–335.
Weiskrantz, L. (1988). *Thought without language.* Clarendon Press. Retrieved from `https://www.harvardlds.org/wp-content/uploads/2017/01/the-origins-of-physical-knowledge-1.pdf`
Wu, J., Yildirim, I., Lim, J. J., Freeman, B., & Tenenbaum, J. (2015). Galileo: Perceiving physical object properties by integrating a physics engine with deep learning. In *Advances in neural information processing systems* (pp. 127–135).