Autonomous Cumulative Transfer Learning

Arash Sheikhlar[1], Kristinn R. Thórisson[1,2], Leonard Eberding[1,3]

[1] Center for Analysis and Design of Intelligent Agents, Reykjavik University, Iceland
arashs@ru.is, thorisson@ru.is
http://cadia.ru.is/
[2] Icelandic Institute for Intelligent Machines, Reykjavik, Iceland
http://www.iiim.is/
[3] Institute of Photogrammetry and GeoInformation,
Leibniz University, Hannover, Germany
l.eberding@stud.uni-hannover.de
https://www.ipi.uni-hannover.de/

Abstract. Autonomous knowledge transfer from a known task to a new one requires discovering task similarities and knowledge generalization without the help of a designer or teacher. How transfer mechanisms in such learning may work is still an open question. Transfer of knowledge makes most sense for learners that must learn numerous novel things, since the value of the ability to transfer will rise with increased experience (other things being equal). In this case a *cumulative learning mechanism* incrementally unifies new knowledge of novel phenomena with existing knowledge, increasing its breadth, depth, and accuracy over time as experience accumulates. Here we address the requirements for what we refer to as autonomous cumulative transfer learning (ACTL) in novel task-environments, including implementation and evaluation criteria, and how it relies on the process of *similarity*, *analogies* and *ampliative reasoning*. While the analysis here is theoretical, the fundamental principles of the cumulative learning mechanism in our theory have been implemented and evaluated in a running system described priorly. We present arguments for the theory from an empirical as well as analytical viewpoint.

Keywords: Transfer Learning · Autonomy · Novelty · Similarity · Ampliative Reasoning · Analogy · Cumulative Learning

1 Introduction

Any agent with general intelligence must be able to deal with novel situations. Since novelty is relative to a learner's knowledge of the world, one way to handle novelty – whether it is a novel juxtaposition of familiar things, a never-before-seen variable or factor, or something entirely new – is to use for guidance priorly experienced situations that seem similar. This is what the canonical concept in psychology of transfer learning (TL) (or *transfer of training*) refers to [5]. What is at stake is an application of prior knowledge and training to a new instance which may (or may not) be mostly identical, or wildly different from what the agent has seen to date. Since novelty is abundant in the physical world, this must be (partly) how the learning process works in many animals: They can *autonomously* (without direct teaching) transfer prior experience to a new situation to (a) classify it, (b) identify its principal factors, in light of active goals, (c) view it in light of prior experience of similar situations, (d) create and initiate the goal-driven actions, in light of currently active goals, and (e) monitor progress in light of predicted outcomes and adjust actions accordingly, possibly involving *a, b, c* and *d*.

This kind of TL requires methods for measures of similarity and relevance, and a compositional knowledge representation. It is cumulative due to the integration of new with old information, but also in that the selective application of prior experience is furthermore subject to learning: If incorrect conclusions are drawn when judging similarity and relevance this can be retroactively dissected, inspected, and learned from. A major mechanism for the comparison is analogies, and this is in turn what is learned: Improved analogy making.

Other kinds of reasoning, however, are necessary also – abduction, deduction, and induction – which means we are really talking about *ampliative reasoning*. The more domain-independent the cumulative learning is, the more effective and efficient knowledge accumulation can become, and this is where ampliative reasoning enters the picture: Using (a) deduction for prediction, based on learned (hypothesized) principles, (b) abduction for deriving plausible causes, and (c) analogies for adapting acquired knowledge to new situations, multiple lines of reasoning can help the learner exclude certain things while highlighting others, more quickly getting to the crux of how to achieve any task in light of prior experience. Finally, (d) induction enables generalization based on invariants across multiple tasks and situations. Reasoning in the physical world must be non-axiomatic because there is no ultimate guarantee that anything is as it seems, and this means that cognitive reasoning cannot follow the rules of formal reasoning [12]. Logic can steer the knowledge accumulation process and enables the cognitive system to make predictions, do planning and transfer its knowledge.

Just as knowledge can take various forms, learning – the effective accumulation of knowledge for future usage – may also rely on various methods. Since each type of situation/task has its own complexities that may make learning and performing difficult for a specific learner, a general machine intelligence (GMI) must be able to not only use the relevant knowledge (knowledge transfer), but also change its own learning style (transfer of learning), based on the features of a particular situation/task. Transfer of learning calls for multiple meta-levels of learning that allows an autonomous cognitive system to choose a knowledge acquisition paradigm that fits a particular task/situation, something which remains to be properly addressed.

To build an intelligent machine that can create its own knowledge and autonomously transfer it to different related situations, evaluate the outcome, and learn from this, all autonomously, an architecture is needed that can make analogies on its own accord and, rather than relying on a human's intuition about similarity and relevance, create its own knowledge for how to do that, based on its own understanding of the world. For GMI the focus needs to be on the actions listed above (a to e), and these need to be integrated with ampliative reasoning. To our knowledge, two approaches to cumulative transfer learning have been demonstrated to date, the AERA system [7] and NARS [13], but an analysis of these with respect to TL remains to be done.

In this paper we present a theoretical analysis of transfer learning (TL), based in part on this prior work, with an attempt to put it into the context of both narrow artificial intelligence (AI) systems and general machine intelligence (GMI). To clearly separate our work on TL from work outside of GMI, we use the term *autonomous cumulative transfer learning* (ACTL). Faced with a novel situation, a learner capable of ACTL selects a specific model or modeling paradigm on its own accord, in light of an analogy that it has itself come up with, creates a new set of models, and uses them in the novel situation. The ACTL process relies in part on selective comparison of similarity and high-level analogy-making. After related work, we describe our theory of autonomous cumulative transfer learning. This is subsequently supported by analytical arguments and empirical data.

2 Related Work

TL has made an appearance in various machine learning (ML) paradigms to date, invariably with the shared goal of increasing learning rate and improving its flexibility. Working on deep neural networks (DNNs) some have implemented a scheme where a human programmer selects a subset of trained network's layers and reuses them to train another network in a similar related domain [14], [15]. Working with the concept of TL in reinforcement learning (RL) techniques others have trained RL for a task and then repurposed it to a similar one [10]. Often in these approaches a human software developer is needed for choosing tasks and making the necessary analogies between the two tasks. Such an approach falls short of what is needed for general machine intelligence (GMI), where the machine must do this automatically and on its own accord, including making the analogies, learning from them, and unifying any new knowledge produced this way with existing knowledge in an explainable manner.

To achieve a positive TL, the agent must determine "what, when, how and why" a knowledge in the memory has to be chosen and transferred to another task. Most of current TL methods assume that the transfer is done offline, that is, happening before the agent starts learning the target task, and thus the question of "when" to transfer has not been much addressed in ML research to date. In addition, a human programmer decides "what" should be transferred, according to some intuition or sense of similarity. Thus, task similarity is another topic that has largely been out of scope in ML research, although a handful of papers have proposed task mappings via concepts from bisimulation [3] and homomorphism [9]. Efficient methodologies are needed to autonomously find similarities that are independent of domain knowledge, although teaching an AI about the domains and the way it can decompose them into sub-domains based on features could be promising [1]. To have an autonomous life-long learning architecture, the learner must be able to find inter-task mappings on its own, do TL while it is learning and performing the target task, and choose the most appropriate knowledge in a specific situation with respect to the time.

One way of thinking about TL is that the knowledge should be created, converted into a generalized form, and stored in the memory in a way that it will be applicable in many future situations. From this point of view, knowledge transfer is, in fact, using the most relevant existing knowledge in different scenarios, in an autonomous manner. This is very similar to how NARS [13] and AERA [8] do TL.

3 Autonomous Cumulative Transfer Learning—A Theory

We consider the *novelty* of an experienced phenomenon Φ a measure on its *familiarity to a cognitive agent*—how similar Φ's aspects are to the agent's available current knowledge. We assume a phenomenon Φ such as state, a process, an occurrence, etc. to consist of aspects[4] made up of elements $\{\varphi_1 \ldots \varphi_n \in \Phi\}$ of various kinds, including relations $\Re_\Phi$ (causal, mereological, sub-structural, etc.) and transitions T_Φ (component processes, transformations, etc., i.e. sub-divisions of Φ), that couple sub-parts of Φ with each other (and with those of other phenomena). Operationally, given a cognitive agent in a task-environment TE and a particular such target phenomenon Φ, if the agent can predict a particular selected aspect $\varphi_n \in \Phi$ using its prior knowledge, then φ_n is familiar to the agent, and non-novel. If the agent can do so for all important aspects of Φ it may be claimed "totally" familiar and non-novel.[5] An agent whose knowledge is compositional – that is, consisting of models made up of smaller models, and can be meaningfully decomposed in a multitude of ways – can, for every complex Φ with a large number of aspects, test its ability to predict each of those aspects $(\varphi_1, \varphi_2 \ldots \varphi_n \in \Phi)$ and record the result; the outcome would be a single number Φ_{nov} that implies a percentage or equivalent of Φ that is novel to an agent, i.e. those aspects the agent failed to predict.

Prediction of a phenomenon must cover the dynamic interference or perturbation by the agent (or something else), and thus some of the relations modelled must include causality. Since causal relations in a lawful world, and in fact all types of relations, make it possible to generate and use rules, the agent's cognitive system must contain some rule-handling mechanisms—reasoning. Reasoning may also be important for selecting which aspects are important for which situations or tasks. Building up knowledge incrementally over time means making

[4] We use 'aspects' as shorthand for 'sub-divisions of a phenomenon that are of *pragmatic importance* to an agent's goals and tasks'.

[5] Since phenomena in the physical world contain an infinite set of subdivisions such a claim would always be limited by pragmatic considerations (see prior footnote). Time and energy will also present hard limits for any such consideration. Thus, there is no literal sense in which complete familiarity may be reached.

a model composed of smaller models that increasingly explains target phenomena, not unlike the process of scientific empirical research. As we have argued elsewhere [11], ampliative reasoning (combined deduction, abduction, induction and analogies) is a way to manage knowledge created under these requirements.

To compute the familiarity of φ the agent must retrieve *relevant* knowledge from its knowledge base, $k \subset KB$, for comparison, $\Phi_{Sim} = \Psi(k, \varphi)$, where Ψ is a multi-dimensional comparison computation using ampliative reasoning and $\Phi_{Sim} \setminus \Phi_{nov}$ is the similarity ratio, relative to the set of target aspects of Φ. Other things being equal, the less familiar something is to the agent, the more novel it is. Novelty is thus always relative to the agent's own current knowledge and is multi-dimensional and continuous.

The transfer learning mechanism we propose states that this multidimensional similarity computation is used to identify overlap between new patterns and previously learned patterns and use it to solve the new task. An autonomous cumulative transfer learner (ACTL) makes analogies and comparisons regarding the number, values, dynamics, and importance of sets of percepts[6] and inferred relations through ampliative reasoning, extracting the importance of each identified perceived variable (or state) and in parallel, discovers the proper set of important variables (of each state), the values of these variables and the dynamics of those values.

The AERA-S1 system [8] was constructed based on a proto-version of this theory and demonstrated to be capable of learning highly complex tasks from observation. Its operational results concur with the theory's predictions, lending the theory some credence. Below we provide further arguments in support for it, from two angles, one analytical the other empirical.

4 Argument from Similarity

One hypothesis we can draw from our theory as outlined above involves percepts. Assuming that a phenomenon at time t is composed of any number of aspects, as defined above. The values of the variables that comprise each aspect can change over time, as well as the precept of the aspect, in accordance with the nature of the phenomenon Φ under observation. The agent can affect aspects of Φ via its actuators; we call variables that can be changed in this way "controllable," V_C. The following is involved in our argument from similarity:

- The available variables V_A. A variable of V_A which can be measured at a time t is an observable of the phenomenon Φ. It should be noted that only a part of V_A might be always observable, and those are the variables that has measurable values in only particular time periods.
- Transition functions (via physical forces).
- prediction functions

[6] The term "percept" as used here references sets of variables in the preceding sense, whether generated by sensors here-and-now, retrieved from memory, or imaginatively constructed.

- The commands the agent produces via its actuators
$$CMD = \{cmd_1(t), cmd_2(t), \ldots, cmd_n(t)\}.$$
which have permissible ranges of values, in accordance with the actuators' specifications.
- The controllable variables the commands can affect,
$$V_c = \{v_{c,1}(t), \ldots, v_{c,n}(t)\}.$$
- The agent's knowledge of a task with respect to the task's goal(s), including all its essential constraints and initial information sufficient to successfully perform the task (and originally bootstrap learning it).
- Knowledge-based similarity estimation functions, *KBSF*, which check values and occurrence times in which causation happened, of available variables inside their observability range. For a causal link, a *SF* finds similarity between the values and the occurrence times via a time and a value threshold.

Our argument rests on comparing states. Similarity functions are defined for variables, states, relations, and transformation functions. The state of a phenomenon is composed of a set of variables which are available for the agent and can be observed and manipulated. A phenomenon going trough changes is considered as a sequence of states, as specified by a set of relevant transition functions (dictated by the world). To compute the similarities between these and relevant knowledge, the variables of the states must be either (already) observed or currently observable by the agent. Taking actions on manipulatable variables produced new states and new aspects of a phenomenon; this is an important method to test predictions, and the process may produce new knowledge about a phenomenon's aspects.

If two or more states, phenomena, or aspects have a high number of identical variables, over the total set of variables (uninon of all vars under consideration), they have ***state similarity in arity (SSA)***. This form of similarity is time- and value-independent. Assume two states, s_1 and s_2; the relationship s_1 SSA s_2 holds if the ratio of identical variables ($V_{idnt} = s_1 \cap s_2$) to all variables ($V = s_1 \cup s_2$) is above an arbitrary threshold ϵ, that is, if the number of elements in V_{idnt} is n, and the number of elements in V is m, we have

$$0 < \epsilon < \frac{n}{m} < 1 \tag{1}$$

Another dimension that's needed we call the ***state similarity of important variables (SSIV)*** between two states. First, the importance of variables must be determined based on the task goal(s).[7]

Each variable is scored with respect to its (predicted) presence in the causal relations that lead to achieving goals.

Therefore, if two or more states have some common variables and the following ratio is above an arbitrary threshold δ, those two states are considered to

[7] This may be done by e.g. backward-chaining from the goal state to the present state using various assumptions about the TE. However, an adequate demonstration and analysis of this process would require its own paper and will not be discussed here.

have SSIV. The SSIV could be calculated in this way: **ratio of** *the sum of the scores of common variables multiplied by 2* **to** *sum of the scores of all variables.*

$$0 < \delta < \frac{2 \cdot \sum_{i=1}^{m} score(v_{idnt}(i))}{\sum_{i=1}^{n} score(v_{nidnt}(i)) + 2 \cdot \sum_{i=1}^{m} score(v_{idnt}(i))} < 1 \tag{2}$$

where m is the number of identical variables, and n is the number of non-identical variables.

The importance is calculated based on scores the variables acquire over time. The scores are given with respect to the repetition of variables' presence in the path(s) of achieving a specific goal. In AERA [8], the saliency of an input (composed of variables with different values - i.e. state) is compared to the system's models, and if the saliency (overall confidence) is higher than a specific threshold, the input will be considered relevant. Here, we introduce an analogous mechanism, which considers a threshold for each of the proposed dimensions, including the arity, value, importance, and transition dynamics in the states. Two or more states, which contain some common variables, have ***state similarity regarding closeness of variables (SSCV)*** if the shared variables have closeness in either values (Value SSCV) or time in which the variables are measurable (Temporal SSCV). Assume s_1 and s_2 with n identical variables, and $v_i, i = 1, .., n$, is an identical variable of s_1 and s_2. Consequently, $v_i(t')$ and $v_i(t'')$, have Temporal SSCV if

$$t'' - t' < \alpha, \qquad t' < t'' \tag{3}$$

and value SSCV if

$$|v_i(t'') - v_i(t')| < \beta_i, \qquad i = 1, ..., n \tag{4}$$

where α and β_i are arbitrary thresholds, and they can be adjusted by the agent through experience with respect to the TE.

The SSCV approach also holds for single **variables similarity** comparison, and we may have ***value variable similarity regarding closeness (value VSC)*** and ***temporal variable similarity regarding closeness (temporal VSC)***.

In the third category, which is **relational similarity**, the relations between states are compared; here we will stick to causal relations. If state $s_1(t)$ is the cause of state $s_2(t')$, and if state $s_3(t'')$ is the cause of state $s_4(t''')$, we can compare the cause states and effect states of $s_1 \rightarrow s_2$ and $s_3 \rightarrow s_4$. In this comparison, the similarity dimensions of SSA, SSCV, and/or SSIV are examined between pair of causes "s_1 *and* s_3" and the pair of effects "s_2 *and* s_4". If the causes are similar, we have ***state casual relational similarity (SCRS)***, and if the effects are similar, we have ***state effectual relational similarity (SERS)***.

The fourth category, **transitional similarity**, relies not only on state similarity but also on confidence of the causal relations between states. If $s_1(t) \rightarrow s_2(t+t')$ holds, it is required to check if $s_3(t'')$ has SSA, SSCV, and/or SSIV with

$s_1(t)$. If so, a prediction is made to produce $s_4(t" + t''')$. Finally, if $s_4(t" + t''')$ has SSA, SSCV, and/or SSIV relation with $s_2(t + t')$, the ***prediction*** is correct and therefore, $s_1(t) \rightarrow s_2(t + t')$ and $s_3(t'') \rightarrow s_4(t'' + t''')$ will have both relational similarity. Then, we can say $s_1(t)$ and $s_3(t'')$ have ***state transitional similarity (STS)***, if both of them with the same transition (or prediction) function reach similar effect states.

If something (variable, state, or relation) is similar to something else (in the same form) in the knowledge base, it is *familiar*. If it turns out that a pair of causal relations has STS, the situation is totally familiar for the agent. Negative knowledge transfer takes place, if the available states/variables have state/variable similarity, while they do not have STS. This makes the agent's predictions fail, since it uses the same improper prediction function for a partially familiar state it observes. Besides, reaching a goal in a task verifies that the agent's predictions were correct. In other words, the agent cannot efficiently reach a goal (except by pure luck) unless it can make acceptably precise predictions. A phenomenon with various aspects may have unforeseen and unpredictable aspects. Therefore, given a phenomenon and agent's knowledge with some similarity to its sensory information, if the agent can make correct predictions about some aspect of the phenomenon, that aspect is totally familiar and therefore not novel.

5 Argument from Empirical Data

We used SAGE [4] to evaluate an actor-critic (AC) reinforcement learner [6] on transfer learning by training it on the cart-pole-task and inverting the forces after initial training. This corresponds to the source task (original) and the target task (inverted). Although in the target task (second phase) the force application has been inverted, all the variables and constraints of the task and observations are the same, and therefore, we have a high SSCV, a high SSIV and a high SSA of the observable states. However, as can be seen in the figure (1), negative transfer learning has happened in the second phase of the learner's lifetime, and re-learning of the target task takes about four times longer, than the original learning of the source task. This means that STS is likely to be very low between the two tasks, since the chosen actions in the source task lead to a negative effect.

This performance degradation stems from the fact that AC learner does not use any metric for autonomous analogy-making during its life time, since task two is, in fact, very similar, and there is only one change in a variable (which, in general, might or might not be important) but drastically reduces the re-learning performance. There is also a third phase in the AC learner's life-time, in which the force has switched back again to the normal mode (after 2500 epochs). As can be seen, again the learner does not find the "force"s importance. Otherwise, it would show faster convergence in the third phase.

This shows that a cumulative learner and of course, an ACTL are needed to spot the novelty in the effect states and compare it to previous effect states identifying cause-effect-chains with high SCRS and low SERS and re-linking the models representing those two causal relations in order to change the way of making prediction and thus, taking actions in the target task.

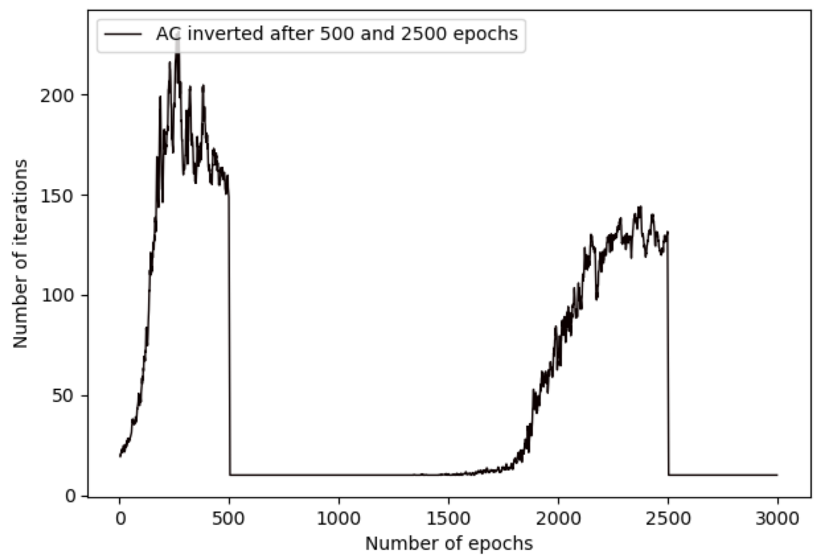

Fig. 1. Training and evaluation on TL of an Actor-Critic (AC) reinforcement learner. Cart-pole task used for training and evaluation derived from [2]. After training period the controllables are inverted (forces from F = [10, -10] N to F = [-10, 10] N). Original training is the source, inverted re-training the target task. AC trained for 500 episodes before inverting. Re-trained for 2000 epsiodes before re-inverting. Evaluated using the SAGE platform [4].

6 Conclusions

In this paper we have introduced a new theory about autonomous cumulative transfer learning (ACTL). It uses similarity measures to identify relevant knowledge in order to transfer it to novel situations during the learner's life-time. This similarity computation relies on analogies, performed in an intertwined manner with non-axiomatic reasoning, which are then used to guide the similarity measurement of a cumulative learner. Similarity as a multidimensional metric to compare situations not only with previously reached states, but rather on the level of states including their composing variables opens the door for further investigation of phenomenon description. Thus this approach might not only be helpful in order to make life-long, cumulative learning possible, but might also give further insights into how a learner can put experience into contexts and domains.

References

1. Jordi E Bieger and Kristinn R Thórisson. Task analysis for teaching cumulative learners. In *International Conference on Artificial General Intelligence*, pages 21–31. Springer, 2018.
2. Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.
3. Pablo Samuel Castro and Doina Precup. Automatic construction of temporally extended actions for mdps using bisimulation metrics. In *European Workshop on Reinforcement Learning*, pages 140–152. Springer, 2011.
4. Leonard Eberding, Arash Sheikhlar, and Kristinn R Thórisson. Sage: Task-environment platform for autonomy and generality evaluation. In *International Conference on Artificial General Intelligence*. Springer, submitted in 2020.
5. Victor Kaptelinin and Bonnie A Nardi. *Acting with technology: Activity theory and interaction design*. MIT press, 2006.
6. Vijay R Konda and John N Tsitsiklis. Actor-critic algorithms. In *Advances in neural information processing systems*, pages 1008–1014, 2000.
7. Eric Nivel and Kristinn R Thórisson. Towards a programming paradigm for control systems with high levels of existential autonomy. In *International Conference on Artificial General Intelligence*, pages 78–87. Springer, 2013.
8. Eric Nivel, Kristinn R Thórisson, Bas R Steunebrink, Haris Dindo, Giovanni Pezzulo, M Rodriguez, C Hernandez, Dimitri Ognibene, Jürgen Schmidhuber, Ricardo Sanz, et al. Bounded recursive self-improvement. *arXiv preprint arXiv:1312.6764*, 2013.
9. Jonathan Sorg and Satinder Singh. Transfer via soft homomorphisms. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 741–748, 2009.
10. Matthew E Taylor and Peter Stone. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10(Jul):1633–1685, 2009.
11. Kristinn R Thórisson, Jordi Bieger, Xiang Li, and Pei Wang. Cumulative learning. In *International Conference on Artificial General Intelligence*, pages 198–208. Springer, 2019.
12. Pei Wang. *Non-axiomatic reasoning system: Exploring the essence of intelligence*. Citeseer, 1995.
13. Pei Wang. *Rigid flexibility: the logic of intelligence*, volume 34. Springer Science & Business Media, 2006.
14. Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. Transfer learning from deep features for remote sensing and poverty mapping. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
15. Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.