
Collaborative creativity with Monte-Carlo Tree Search and Convolutional Neural Networks

Memo Akten

Department of Computing
Goldsmiths University of London
m.akten@gold.ac.uk

Mick Grierson

Department of Computing
Goldsmiths University of London
m.grierson@gold.ac.uk

Abstract

We investigate a human-machine collaborative drawing environment in which an autonomous agent sketches images while optionally allowing a user to directly influence the agent’s trajectory. We combine Monte Carlo Tree Search with image classifiers and test both shallow models (e.g. multinomial logistic regression) and deep Convolutional Neural Networks (e.g. LeNet, Inception v3). We found that using the shallow model, the agent produces a limited variety of images, which are noticeably recognisable by humans. However, using the deeper models, the agent produces a more diverse range of images, and while the agent remains very confident (99.99%) in having achieved its objective, to humans they mostly resemble unrecognisable ‘random’ noise. We relate this to recent research which also discovered that ‘deep neural networks are easily fooled’ [30] and we discuss possible solutions and future directions for the research.

1 Introduction

The aim of this research is to create an autonomous, collaborative, ‘creative’ agent that can sketch images in realtime, while optionally allowing a user to intervene and physically influence the agent, affecting how and what it decides to draw. To be more specific, our objective is to allow the agent to take random actions, and sketch any path (i.e. follow any trajectory) that resembles a predetermined, desired object *class*. E.g. everytime we ask the agent to draw a ‘cat’, it should draw a different picture of a cat. Note, the objective for the agent is not to sketch an image that resembles a *specific target picture*, as in that case, no matter how the agent meanders, it would always converge back to the specified target picture. Similarly, we stay away from *any number of* target pictures and instead turn to *Deep Learning* to exploit latent representations that might be useful in aiding the agent. We also allow a user to optionally interact with the agent, and exert forces on it to push and pull it around, causing the agent to diverge onto a new trajectory that might lead to a new picture — which still resembles the desired class (e.g. it sketches a different ‘cat’ picture). The output images would all be varied, and in effect be collaborations between the user and the agent. The entire system could be seen as a realtime collaborative drawing environment between a human and a ‘creative’ autonomous agent. This research is presented as work in progress. At this stage we have mixed results (discussed in Section 4) where the system shows potential, but also demonstrates weaknesses relating to the Deep Learning models.

2 Background

Computer generated, procedural visual art is a rich field with its roots dating back to the 1950s with John Whitney Sr., followed by artists such as Paul Brown, Vera Molnar, Manfred Mohr, Frieder Nake, Larry Cuba and more from the 1960s onwards [14]. Harold Cohen’s AARON [8], was

arguably the first software using Artificial Intelligence (AI) to produce visual art, followed by artists such as William Latham [40], Karl Sims, [35], and Scott Draves [11]. Also in the 1960s and 70s, Myron Kruger introduced interactivity — particularly *realtime gestural interactivity* — into computer artworks, culminating in his seminal Artificial Reality *Responsive Environment* ‘Videoplace’ [20]. More recently, works such as Simon Colton’s ‘Painting Fool’ [9] and Patrick Tresset’s ‘Artistically Skilled Embodied Agents’ [41] also investigate the ‘psychology’ of ‘creative’ AI agents.

While this is the context for our research, computational models of creativity — such as those described in [6] — or methods of evaluating machine creativity — such as those described in [32] — are currently not within the scope of this study. Instead, we focus on investigating the opportunities provided by Deep artificial Neural Networks (DNNs) — in particular *Convolutional Neural Networks (CNN)* — combined with agent-based AI methods — such as *Monte Carlo Tree Search (MCTS)* — for the task of driving collaborative drawing agents.

Deep convolutional architectures saw initial success in image classification in the late 1980s [21], and they are now consistently providing state of the art results in this field [19, 38, 16, 37]. An active area of research is focused on trying to understand and visualise the internal workings of CNNs, and use them for generative purposes [42, 25, 26, 33, 28, 12]. Additionally, CNNs have also been used with great success to drive AI agents, such as those playing Atari games, Doom or Go [27, 15, 18, 34]. In these latter cases, the CNNs have effectively acted as the agent’s ‘eyes’, breaking down the raw input (i.e. screen pixels or state of the Go board) into more abstract, meaningful representations which can be used by the agent. These researchers have also used MCTS [7] to guide the agents, and allow them to make more optimal decisions based on the meaningful representations provided by the CNN. It is based on this direction of research, and within the artistic, ‘artificially creative’ context outlined above that our study takes place.

3 Method

The agent, similar to a LOGO Turtle [31], can move forwards (leaving a trail as it moves), or rotate clockwise or counter-clockwise. We use MCTS to simulate many alternative actions and paths at every timestep, and at the end of each simulation, the entire trajectory of the agent — including the previous history as well as the simulation — is fed into an image classifier. The classifier evaluates how much the agent’s trajectory resembles the desired class, and that score is used as the reward backpropagated by MCTS to choose the final action. This is described in more detail in the sections below.

We tested three classifiers trained on two datasets: i) the MNIST dataset of hand-written digits [23], and ii) the ImageNet dataset of millions of labeled images classified into one thousand classes[10].

We developed our application in *openFrameworks* [24], a C++ creative development toolkit very popular amongst artists and creative coders. We used *ofxMSAmcts*, our own open-source C++ MCTS implementation (i.e. *addon*) for *openFrameworks* [2]. We built, trained and saved the image classifiers in python using TensorFlow [1], and loaded the trained models into C++/*openFrameworks* using *ofxMSATensorFlow*, again our own open-source TensorFlow *openFrameworks* wrapper (*addon*) [3].

3.1 MCTS Agent

The MCTS procedure is summarised in Appendix A and a more detailed survey can be found in [7]. In our implementation, our agent can move forward with a number of different speeds (n_{speeds}) to choose from, ranging from s_{min} to s_{max} . It can rotate in r_{inc} increments ranging from $-r_{max}$ to $+r_{max}$, giving it $n_{rot} = 2r_{max}/r_{inc}$ number of different rotations to choose from. Thus at each timestep, the agent has $n_a = n_{speeds} * n_{rot}$ total number of actions to choose from. An example configuration is: $n_{speeds} := 2, s_{min} := 0, s_{max} := 2, n_{rot} := 7, r_{max} := 30,$ with $n_a = 14$.

The state s at any timestep, is the full trajectory (stored as a vector path) that the agent has followed (i.e. sketched) up to that point in time. The state s' which is reached after the agent takes an action a , is the full image that will be drawn once the agent takes that action. The MCTS simulations end after a maximum rollout depth (e.g. $D = 100$), after which the final simulated drawn image (rasterized as a bitmap image) is fed into an image classifier, and the probability score for the desired class is retrieved and backpropagated as the reward for reaching that state. In our application we also visualise all of the ‘imagined’ paths (i.e. all simulations).

3.2 MNIST - Multinomial logistic regression

The first classifier we train is a *multinomial logistic regression* trained on the MNIST dataset, a linear transformation followed by a softmax activation [5]. We formulate this as

$$\mathbf{y} = \text{softmax}(\mathbf{W}\mathbf{x} + \mathbf{b}), \quad \text{where} \quad \text{softmax}(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}}, \quad (1)$$

\mathbf{x} is a vector containing the flattened input image pixels, and \mathbf{W} and \mathbf{b} are respectively the *weights matrix* and *bias vector* to be learnt through training. We train using stochastic gradient descent and backpropagation trying to minimise the cross-entropy [5] between \mathbf{y} , the predicted probability distribution of our model, and \mathbf{y}' , the true distribution, i.e. the training data.

$$H(\mathbf{y}', \mathbf{y}) = - \sum_i y'_i \log(y_i) \quad (2)$$

The model completed training in a few seconds (on CPU) and unsurprisingly it doesn't generalise very well, scoring about 90% on the validation data. As expected, it is very restrictive on the types of inputs it can classify and does not provide much translational, rotational or scale invariance as the softmax operates directly on a linear transformation of the flattened raw pixel data.

3.3 MNIST - Convolutional Neural Network (LeNet)

The second classifier we train is a deep Convolutional Neural Network (CNN) similar to the LeNet architecture [22]. CNNs are deep architectures inspired by the visual cortex, and consist of a hierarchy of *stacked feature maps*, where each layer extracts features from the layer below using two-dimensional convolution kernels which are also learnt during training. We use two convolutional stacks, each stack consisting of 5x5 convolution kernels followed by an element-wise rectified linear unit ($ReLU(x) = \max(0, x)$ [17]), followed by a 2x2 max-pool layer. The first convolutional stack learns 32 feature maps while the second stack learns 64 feature maps. After the two convolution stacks, we use a dense fully connected layer with 1024 neurons, followed by a softmax to convert the results to a normalised probability distribution. Similar to the multinomial logistic regression, we train with stochastic gradient descent and backpropagation to minimise the cross-entropy. We use dropout regularization with a probability of 50% [36].

The model trained in a couple of minutes (again on CPU). It scored 99.2% accuracy on the validation data and unsurprisingly proved to be much more successful in classifying images, with higher resilience to noise, translation, scale and rotation compared to the multinomial logistic regression.

3.4 ImageNet - Inception-v3

For the final classifier, we download and use a pre-trained model of Google's state of the art image classification architecture Inception-v3 [39], trained on ImageNet. This is a deep CNN reaching as low as 3.58% error in the 2012 ImageNet validation set for Top-5 error.

4 Results and conclusion

The overall system architecture seems to work and shows potential.

Using MNIST with the shallow model — multinomial logistic regression — proved very effective and positive results were obtained. Example output can be seen in Fig. 1 (a), where the agent has clearly sketched the desired image which was the digit '3'. Running this multiple times with different objectives provided similar results, demonstrating that the MCTS + Image Classifier system works on the whole.

However, we found that the deep CNNs were not ideal for generative applications in this manner, as they provide too many false positives. As discriminative models trained to classify images, they can classify *natural* images *correctly* with very high accuracy and confidence, but they also *incorrectly* classify *unnatural* images such as noise and abstract shapes, with equally high confidence. In

other words, the class manifolds successfully represent the desired *natural images*, but also include *undesired unnatural images* as well. In retrospect this is not too surprising, as these networks are designed to function under very high noise with high degree of translational, rotational and scale invariance, and the training data does not contain any unnatural images.

Ultimately, the tendency of these deep models to identify false positives, cause the agent to wander around the screen randomly, sketching out images that appear to humans as unrecognisable ‘random’ noise, but are classified by the network with very high (99.99%) confidence to be the desired class. This can be seen in Fig. 1 (b) and (c). Interestingly, even though the system produces what appears to humans as random noise, it’s not only the *top* prediction of the network which classifies the outputs as the desired class, but the *top 5* predictions as well. E.g. in Fig. 1 (b) the desired class is ‘meerkat’. The top 5 predictions of the network are ‘meerkat’ (99.999%), ‘mongoose’, ‘hyena’, ‘Egyptian cat’, ‘cheetah’ — all small, cat-like animals. In Fig. 1 (c) the desired class is ‘white wolf’. The top 5 predictions are ‘white wolf’ (97.8%), ‘timber wolf’, ‘Arctic fox’, ‘Samoyed’, ‘west highlands white terrier’ — all white/light grey dogs.

In summary, the failure of the system — i.e. the output not resembling the desired class in our eyes — is not a failure of the *whole* system, or of MCTS, or the way MCTS and the classifier are integrated. But it is a failure of the deep CNN classifier, in that it’s easily susceptible to false positives. This also confirms recent research that deep CNNs can easily be fooled [30]. Though this previous research demonstrated this using evolutionary algorithms, and in a very non-realtime, non-interactive manner, as opposed to our realtime agent based methods. There are also recent examples of overcoming this problem of generating *unnatural images*, by using *natural image priors* [28, 29] or adversarial networks [13, 33]. Both approaches we are planning to try in future research.

At every timestep, we also visualise all of the MCTS rollouts (Fig. 3). From a conceptual and aesthetic point of view this can be thought of as visualising the possible trajectories ‘imagined’ by the agent, which is in part the motivation for this research.

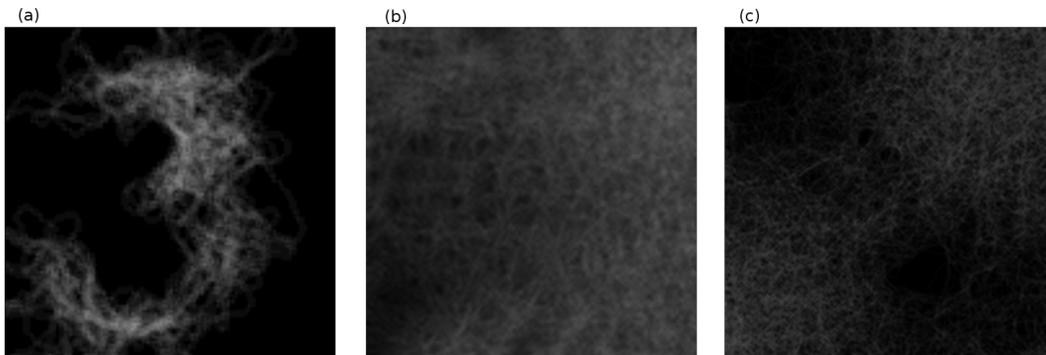


Figure 1: Agent trying to sketch (a) the digit ‘3’ using Multinomial Logistic Regression trained on MNIST as the classifier, (b) a ‘meerkat’ using Google’s pre-trained state of the art Inception-v3 model as image classifier, (c) a ‘wolf’ using the same system. Interestingly, the output of the latter two looks like random noise to us, but the classifier evaluates both images with extremely high confidence to be the desired class. See Fig. 3 for full screenshot including confidence values and simulations.

References

- [1] M. Abadi and A. Others. TensorFlow : Large-Scale Machine Learning on Heterogeneous Distributed Systems, 2015.
- [2] M. Akten. ofxMSAmcts. 2015.
- [3] M. Akten. ofxMSATensorFlow, 2016.
- [4] R. Bellman. A Markovian decision process. Technical report, DTIC Document, 1957.
- [5] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.
- [6] M. a. Boden. Creativity and artificial intelligence. *Artificial Intelligence*, 103(1-2):347–356, 1998.

- [7] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1):1–43, 2012.
- [8] H. Cohen. Parallel to perception: some notes on the problem of machine-generated art. *Computer Studies*, pages 1–10, 1973.
- [9] S. Colton. The painting fool: Stories from building an automated painter. In *Computers and creativity*, pages 3–38. Springer, 2012.
- [10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009.*, pages 248–255. IEEE, 2009.
- [11] S. Draves. The Electric Sheep screen-saver: A case study in aesthetic evolution. *Proc. EvoMUSART*, pages 458–467, 2005.
- [12] L. a. Gatys, A. S. Ecker, and M. Bethge. A Neural Algorithm of Artistic Style. *arXiv preprint arXiv:1508.06576*, 2015.
- [13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.
- [14] M. Grierson. *Audiovisual composition*. PhD thesis, 2005.
- [15] X. Guo, S. Singh, H. Lee, R. Lewis, and X. Wang. Deep Learning for Real-Time Atari Game Play Using Offline Monte-Carlo Tree Search Planning. *Advances in Neural Information Processing Systems (NIPS)*, 2600:3338–3346, 2014.
- [16] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. *arXiv preprint arXiv:1512.03385*, 2015.
- [17] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. Lecun. What is the best multi-stage architecture for object recognition? pages 2146–2153. IEEE, 2009.
- [18] M. Kempka, M. Wydmuch, G. Runc, J. Toczek, and W. Jaśkowski. ViZDoom: A Doom-based AI Research Platform for Visual Reinforcement Learning. *arXiv:1605.02097v1 [cs.LG]*, 2016.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems (NIPS)*, pages 1–9, 2012.
- [20] M. W. Krueger, T. Gionfriddo, and K. Hinrichsen. VIDEOPLACE—an artificial reality. *ACM SIGCHI Bulletin*, 16(4):35–40, 1985.
- [21] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.
- [22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2323, 1998.
- [23] Y. LeCun and C. Cortes. MNIST handwritten digit database. *AT&T Labs [Online]*. Available: <http://yann.lecun.com/exdb/mnist>, 2010.
- [24] Z. Lieberman, T. Watson, and A. Castro. OpenFrameworks, 2016.
- [25] A. Mahendran and A. Vedaldi. Understanding Deep Image Representations by Inverting Them. 2014.
- [26] A. Mahendran and A. Vedaldi. Visualizing Deep Convolutional Neural Networks Using Natural Pre-Images. 2015.
- [27] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing Atari with Deep Reinforcement Learning. *arXiv preprint arXiv: . . .*, pages 1–9, 2013.
- [28] A. Mordvintsev, C. Olah, and M. Tyka. Deepdream inceptionism, 2015.
- [29] A. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox, and J. Clune. Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. *Advances in Neural Information Processing Systems 29*, 2016.
- [30] A. Nguyen, J. Yosinski, and J. Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 427–436, 2015.
- [31] S. Papert, J. A. Valente, and B. Bitelman. Logo: computadores e educa, 1980.
- [32] A. Pease, D. Winterstein, S. Colton, A. Pease, D. Winterstein, and S. Colton. Division of Informatics , University of Edinburgh Centre for Intelligent Systems and their Applications by Evaluating Machine Creativity. *Workshop on Creative Systems, 4th International Conference on Case Based Reasoning*, (November):129–137, 2001.

- [33] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [34] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. V. D. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.
- [35] K. Sims. Evolving virtual creatures. *Siggraph '94, SIGGRAPH '(July):15–22*, 1994.
- [36] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research (JMLR)*, 15(1):1929–1958, 2014.
- [37] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv preprint arXiv:1602.07261*, 2016.
- [38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception Architecture for Computer Vision. *arXiv preprint arXiv:1512.00567*, 2015.
- [40] S. Todd and W. Latham. *Evolutionary art and computers*. Academic Press, Inc., 1992.
- [41] P. Tresset and O. Deussen. Artistically Skilled Embodied Agents. (April), 2014.
- [42] M. D. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. *arXiv preprint arXiv:1311.2901*, 2013.

A Monte Carlo Tree Search

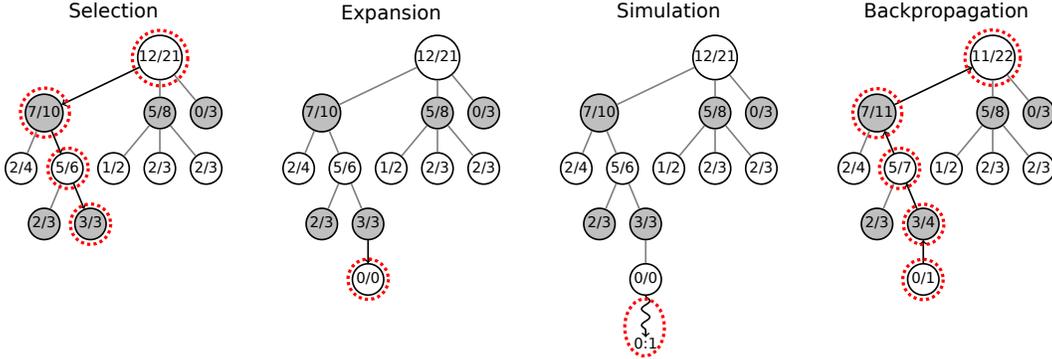


Figure 2: Overview of MCTS. Image from wikipedia, CC BY-SA 3.0

MCTS is a probabilistic approach to searching a decision tree. At its core, it treats actions like slot machines in a multi-armed bandit. Below we briefly discuss our implementation. A detailed explanation and survey can be found in [7].

We represent our problem as a discrete-time *Markov Decision Process* (MDP) [4]. At a particular timestep the system is in a *state* s . The agent can take an *action* a (out of a number of available actions n_a) to arrive at a new state s' , and will be given a *reward* $R_a(s, s')$ for doing so. Instead of trying to simulate every possible action at every timestep, MCTS uses a heuristic to only partially expand what might be the most promising branches of the decision tree. At every timestep, many simulations are run, but only from carefully chosen nodes of the decision tree. The algorithm decides which node to expand and runs a simulation from it based on a balance between *exploiting* known rewards for nodes already fully expanded, vs *exploring* nodes which have not yet been fully explored, i.e. with unknown rewards.

An overview of MCTS can be seen in Fig. 2 and summarised as i) *Select* a node to expand by starting at the current root node and recursively stepping through fully expanded child nodes until an *expandable node* is reached (i.e. a non-terminal node with unvisited children), ii) *Expand* the selected node (i.e. pick one of its unvisited children), iii) Run a *simulation* (i.e. *rollout*) from the new child node until an *outcome* is reached, iv) *Backpropagate* a reward value Δ based on the outcome of the simulation, back up the tree, accumulating statistics for each node.

At each timestep, this procedure of $\{selection-expansion-simulation-backpropagation\}$ is repeated many times to partially expand the decision tree, until a predefined criterion is met. This is usually an allocated time budget (e.g. maximum milliseconds per timestep) or an upper bound on the number of simulations per timestep, or a domain-specific interruption criterion. Once the search is terminated, a final decision is made by choosing an action according to a criterion based on the search statistics. E.g. The most visited child node (*Robust Child*) or child with highest rewards (*Max Child*), more examples can be found in [7].

The *tree policy* decides which node to select and expand, and we use the *Upper Confidence Bound 1 applied to Trees* (UCT), formulated as

$$\frac{v}{n_c} + k\sqrt{\frac{2\ln(n_t)}{n_c}} \quad (3)$$

where v is the *value* of the child node, n_c is the number of times the child node has been visited, k is the *exploration parameter* (theoretically equal to $\sqrt{2}$ but usually chosen empirically) and n_t is the total number of simulations for the node considered. The *default policy* decides how to choose actions during the simulation (i.e. rollout). We use a *random rollout*, i.e. actions are selected randomly from the n_a available actions.

In a typical two-player zero-sum game such as tic-tac-toe, MCTS simulations will run until they reach a *terminal state*, where the outcome is either a *win*, *lose*, or *draw*. In these situations, usually a simple reward value $\Delta = 1$ is backpropagated if the agent wins, $\Delta = -1$ if the agent loses, and $\Delta = 0$ if the result is a draw. Since each simulation iteration combinatorially increases the size of the decision tree, in situations where the tree is very *deep* — such as in the game of Go [34] — it is not feasible to run simulations until a terminal state is reached. In these cases it is common to prematurely end the simulation after a predefined *maximum rollout depth* D , and then evaluate or approximate the reward value Δ using a heuristic on the simulation end-state. This approximated reward value is then backpropagated as before.

B Screenshots

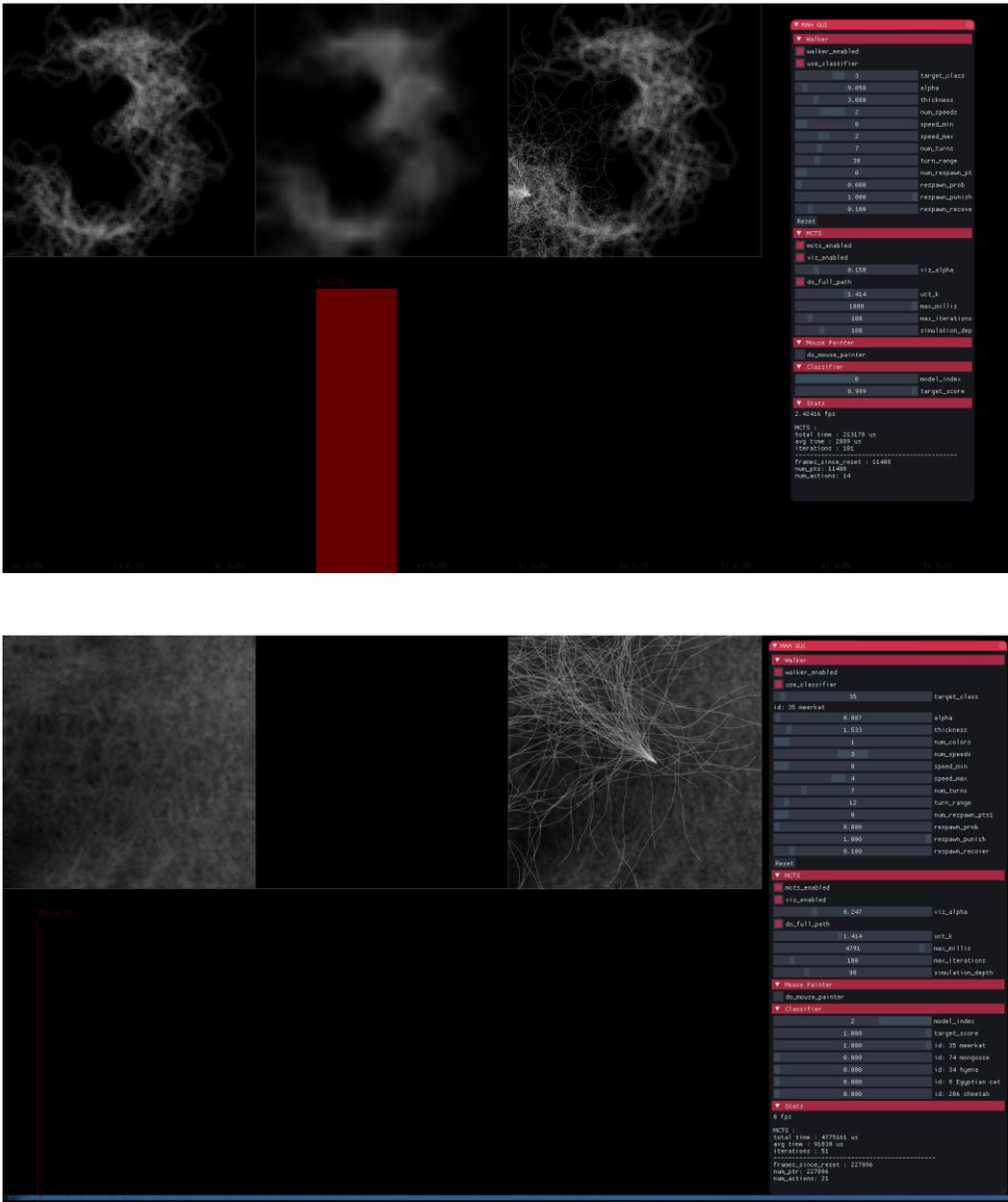


Figure 3: Full screenshots for agent sketching and simulation software. Top: sketching the digit '3' using the Multinomial Logistic Regression trained on MNIST as the classifier, bottom: sketching a 'meerkat' using Google's pre-trained state of the art Inception-v3 model as image classifier. In both instances the top-left viewport shows the current output, top-right viewport shows a visualization of the current MCTS rollouts (showing all simulated paths which will be evaluated by the classifier).