World Models

David Ha, Jürgen Schmidhuber

Edmund Goodman

November 20, 2024

How do we experience the world?

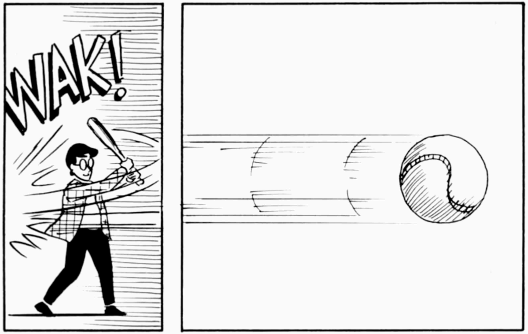

Figure 1: Art by Scott McCloud[a].

- Humans build spatial and temporal models of the environment we experience
  - Sometimes actions occur so fast we work instinctively from these models
  - Predicting rather than processing
- Can we build neural networks which operate similarly?

[a] McCloud and Martin, *Understanding comics: The invisible art*.

Existing work

1990: RNN model-controllers (right)[a]

2012: AlexNet and deep neural networks[b]

2013: Variational auto-encoders[c]

2018: World models[d]

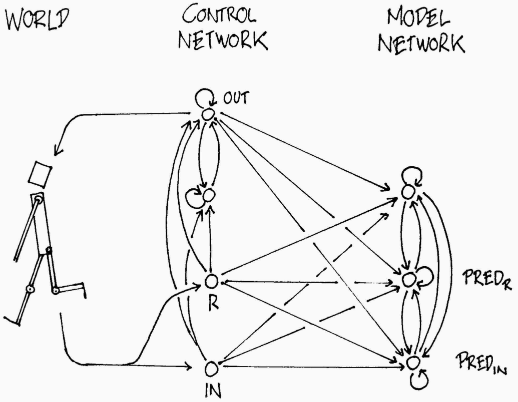

Figure 2: A controller with internal RNN model of the world.

---

[a] Schmidhuber, *Making the world differentiable: on using self supervised fully recurrent neural networks for dynamic reinforcement learning and planning in non-stationary environments*, Figure 2.

[b] Krizhevsky, Sutskever, and Hinton, "ImageNet Classification with Deep Convolutional Neural Networks".

[c] Kingma and Welling, *Auto-Encoding Variational Bayes*.

[d] Ha and Schmidhuber, *World Models*.

Key question and contributions

"Can agents learn inside of their own dreams?"[1]

- Combine existing approaches (model-controller RNNs, DNNs, variational auto-encoders) into state-of-the-art generative models for game environments
- Show that agents can be trained through the lens of their own generative models (their dreams)

[1] Ha and Schmidhuber, *World Models*.

Components

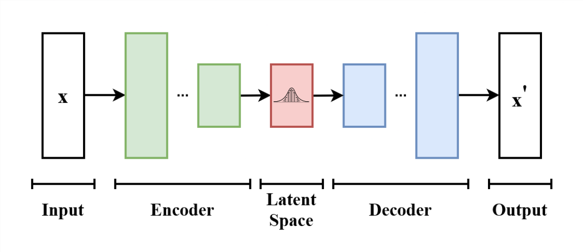

Figure 3: A diagram of a variational auto-encoder[a].

[a] EugenioTL, *Variational Autoencoder structure*.

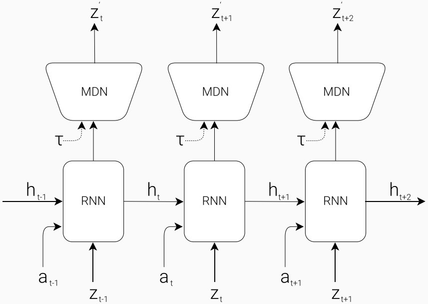

Figure 4: A diagram of an RNN with a mixture density network output layer[a].

[a] Ha and Schmidhuber, *World Models*, Figure 6.

Architecture

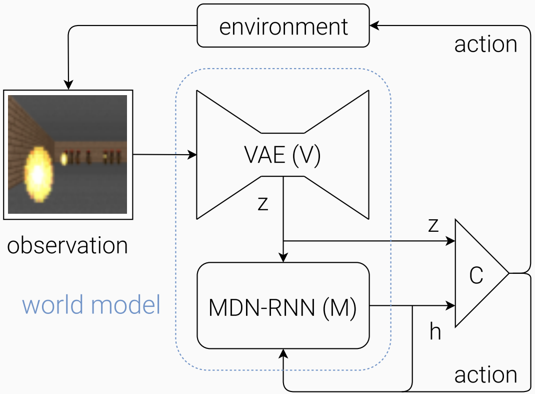

Figure 5: Flow diagram of the agent model[a]

- Three components to model
  - **V:** Learns to represent spatial component of the environment as latent representation z
  - **M:** Learns to predict temporal component of the environment
  - **C:** Learns to maximise reward from world model only
- $V + M$ are the world model – large, but can be trained unsupervised from environment
- C adds agency – small (single-layer), takes features from world model as input

[a] Ha and Schmidhuber, *World Models*, Figure 8.

Training cars to race

Figure 6: A photo[a] of `CarRacing-v0` from OpenAI's gym[b]

1. Collect 10,000 rollouts from a random policy
2. Train VAE (V) to encode frames into $z \in \mathcal{R}^{32}$.
3. Train MDN-RNN (M) to model $\mathbb{P}(z_{t+1}|a_t, z_t, h_t)$.
4. Define Controller (C) as $a_t = W_c \; [z_t \; h_t] + b_c$.
5. Use CMA-ES[a] to solve for a W_c and b_c that maximizes the expected cumulative reward

[a] Ha and Schmidhuber, *World Models*, Figure 11.
[b] *Car Racing - Gym Documentation*.

[a] Loshchilov and Hutter, *CMA-ES for Hyperparameter Optimization of Deep Neural Networks*.

Winning races

Method	Avg. Score
DQN (Prieur, 2017)	343 ± 18
A3C (continuous) (Jang et al., 2017)	591 ± 45
A3C (discrete) (Khan & Elibol, 2016)	652 ± 10
ceobillionaire (Gym Leaderboard)	838 ± 11
V model	632 ± 251
V model with hidden layer	788 ± 141
Full World Model	$\mathbf{906 \pm 21}$

Figure 7: `CarRacing-v0` scores achieved using various methods[2].

- Spatial only $(V + C)$ model is fairly effective, albeit with unstable driving
- Full world $(V + M + C)$ model is best-in-class, “attacking” sharp corners

[2] Ha and Schmidhuber, *World Models*, Table 1.

Do agents dream of electric cars?

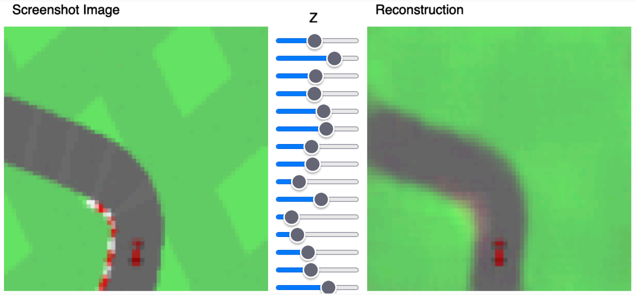

Figure 8: Car racing observation and reconstruction from autoencoder – interactive demo available: `https://worldmodels.github.io/`

- With the trained MDN-RNN, we can predict the next state z_{t+1} from z_t and the action
- What if we used this prediction instead of an empirical observation?

Learning from dreams

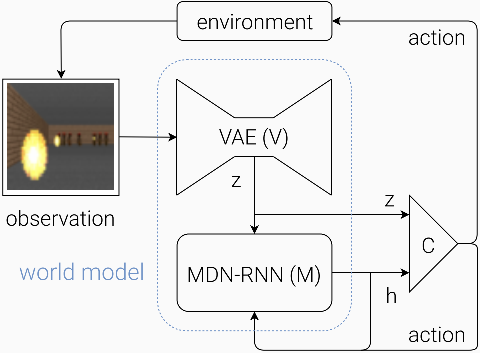

Figure 9: Flow diagram of the agent model[a].

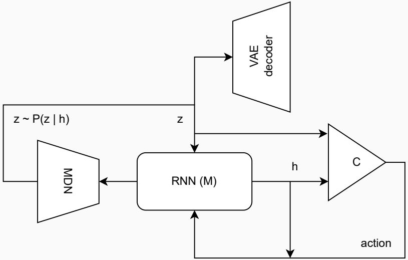

Figure 10: Modified agent model, "learning inside a dream".

[a] Ha and Schmidhuber, *World Models*, Figure 8.

VizDoom experiment

Figure 11: Screenshot of the "VizDoom: Take Cover" environment[a].

- Similar setup to the Car Racing experiment, but this time all learning is done in dreams
- This works! Agents can learn inside their own dreams, with this learnt policy being effective in the actual environment
- There are a few issues:
  - Model doesn't perfectly represent environment, so agent can "cheat", resolved by leveraging temperature
  - Complex environments are hard to search comprehensively, resolved by iteratively training

[a] Ha and Schmidhuber, *World Models*, Figure 14.

Impact

- Influential in the ongoing development of foundation models
  - "The first work that proposes to learn a compressed spatial and temporal representation of the environment in an unsupervised manner using a simple Variational Autoencoder"[3].
- Resulted in the "Dreamer" series of papers by Google DeepMind:
  1. Dreamer solves long-horizon tasks using latent imagination of reinforcement learning[4]
  2. DreamerV2 then uses this approach to successfully play Atari games[5]
  3. DreamerV3 further extends this approach to generally solve tasks without human input[6]

[3] Zhou et al., *A Comprehensive Survey on Pretrained Foundation Models*, Appendix E.
[4] Hafner, Lillicrap, Ba, et al., *Dream to Control*.
[5] Hafner, Lillicrap, Norouzi, et al., *Mastering Atari with Discrete World Models*.
[6] Hafner, Pasukonis, et al., *Mastering Diverse Domains through World Models*.

Criticism and future work

Strengths:

+ Proposes architecture which outperforms existing work on competitive benchmarks
+ Demonstrates that training in dreams learns effective policies

Weaknesses:

– Motivations for training in dreams only mentioned briefly – demonstrations of how it facilitates training without expensive simulation would be helpful
– Reward function separated from spatial/temporal feature extraction, causing unnecessary artefacts
– Approach is "instinctive" – no mechanism for planning far ahead

Future work:

⇒ Including reward function in spatial and temporal models
⇒ Hierarchical models to support planning and strategy

World Models

David Ha, Jürgen Schmidhuber

Edmund Goodman

November 20, 2024

References i

[1] *Car Racing - Gym Documentation.* URL: https://www.gymlibrary.dev/environments/box2d/car_racing/ (visited on 11/19/2024).

[2] EugenioTL. *Variational Autoencoder structure.* July 2021. URL: https://commons.wikimedia.org/wiki/File:VAE_Basic.png#/media/File:VAE_Basic.png (visited on 11/20/2024).

[3] David Ha and Jürgen Schmidhuber. *World Models.* arXiv:1803.10122. May 2018. DOI: 10.48550/arXiv.1803.10122. URL: http://arxiv.org/abs/1803.10122 (visited on 11/17/2024).

[4] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. ***Dream to Control: Learning Behaviors by Latent Imagination.*** arXiv:1912.01603. Mar. 2020. DOI: 10.48550/arXiv.1912.01603. URL: http://arxiv.org/abs/1912.01603 (visited on 11/20/2024).

[5] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. ***Mastering Atari with Discrete World Models.*** arXiv:2010.02193. Feb. 2022. DOI: 10.48550/arXiv.2010.02193. URL: http://arxiv.org/abs/2010.02193 (visited on 11/20/2024).

[6] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. ***Mastering Diverse Domains through World Models.*** arXiv:2301.04104. Apr. 2024. DOI: 10.48550/arXiv.2301.04104. URL: http://arxiv.org/abs/2301.04104 (visited on 11/20/2024).

References iii

[7] Diederik P. Kingma and Max Welling. ***Auto-Encoding Variational Bayes.*** arXiv:1312.6114. Dec. 2022. DOI: 10.48550/arXiv.1312.6114. URL: http://arxiv.org/abs/1312.6114 (visited on 11/19/2024).

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. **"ImageNet Classification with Deep Convolutional Neural Networks".** In: *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc., 2012. URL: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (visited on 11/12/2024).

[9] Ilya Loshchilov and Frank Hutter. ***CMA-ES for Hyperparameter Optimization of Deep Neural Networks.*** arXiv:1604.07269. Apr. 2016. DOI: 10.48550/arXiv.1604.07269. URL: http://arxiv.org/abs/1604.07269 (visited on 11/19/2024).

[10] Scott McCloud and Mark Martin. ***Understanding comics: The invisible art.*** Vol. 106. Kitchen sink press Northampton, MA, 1993.

[11] Jürgen Schmidhuber. ***Making the world differentiable: on using self supervised fully recurrent neural networks for dynamic reinforcement learning and planning in non-stationary environments.*** Vol. 126. Inst. für Informatik, 1990.

[12] Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, Hao Peng, Jianxin Li, Jia Wu, Ziwei Liu, Pengtao Xie, Caiming Xiong, Jian Pei, Philip S. Yu, and Lichao Sun. ***A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT.*** arXiv:2302.09419. May 2023. DOI: `10.48550/arXiv.2302.09419`. URL: `http://arxiv.org/abs/2302.09419` (visited on 11/20/2024).