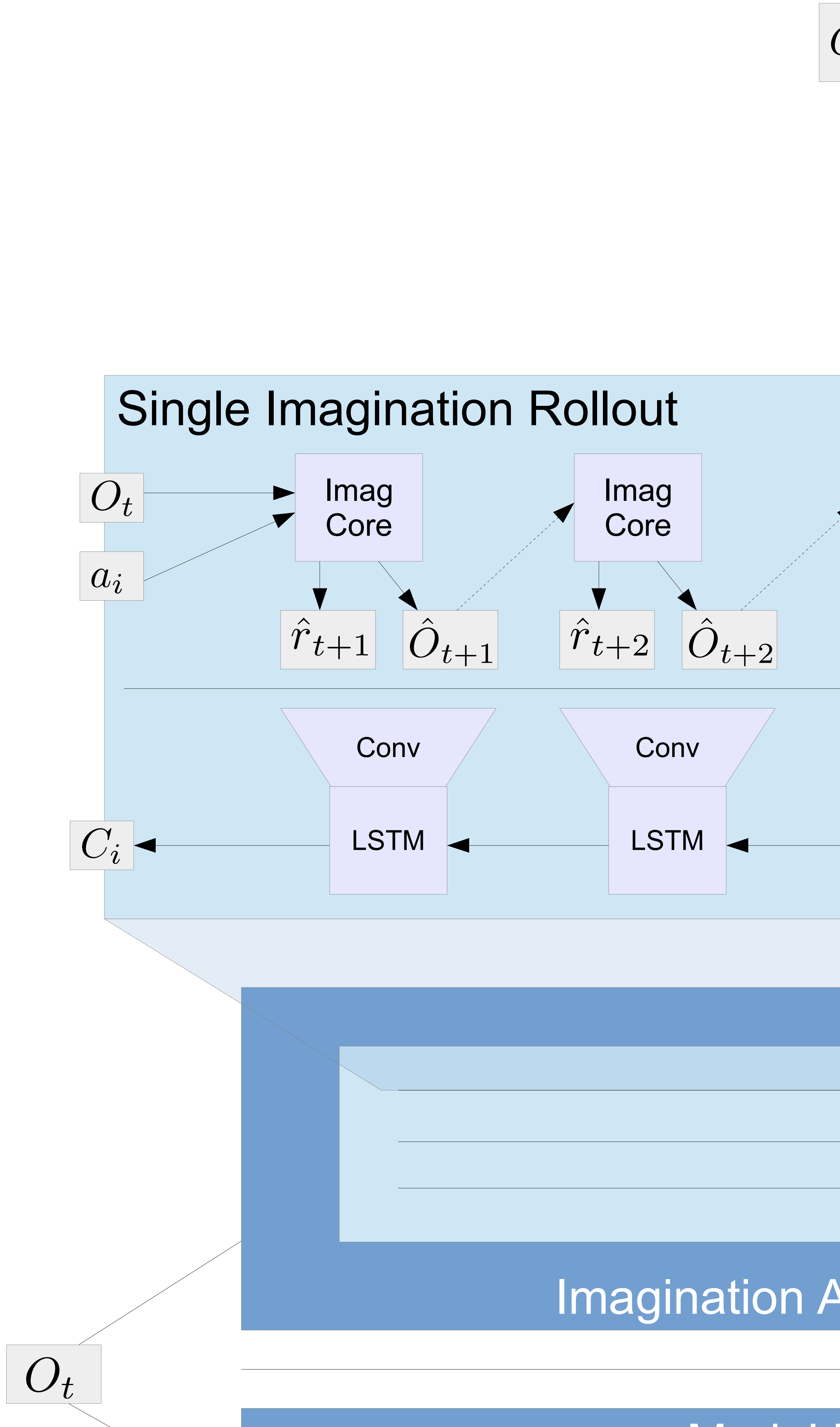
Imagination A
Reinforceme

Angela Denniger, Felix Sch

I

- **Adopted implementation** of the pape *Learning* by DeepMind [1] (I2A)
- We were not able to replicate the resul they used a custom implementation of **environment**. [1,4]
- Combines **model based and model fr**
- Different **Imagination Rollouts explor**

Full I2A Architecture



Single Imagination Rollout
O_t
a_i
Imag Core
Imag Core
$\hat{r}_{t+1}$
$\hat{O}_{t+1}$
$\hat{r}_{t+2}$
$\hat{O}_{t+2}$
Conv
Conv
LSTM
LSTM
C_i
O_t
Imagination A

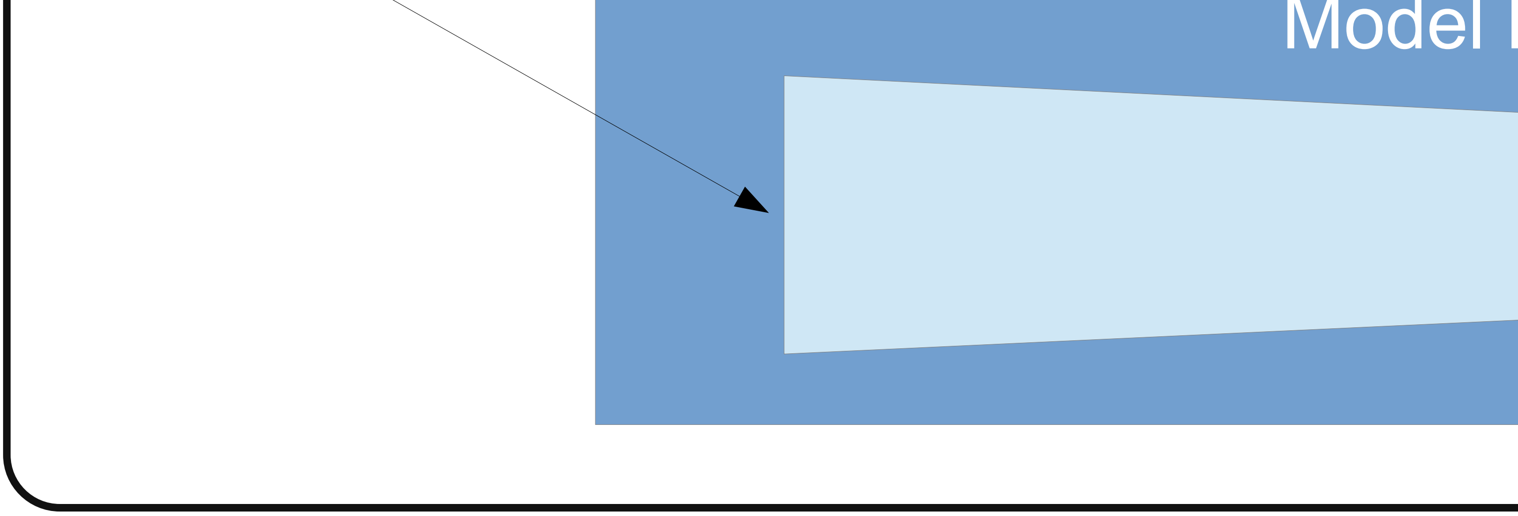


Fig. 1: Full Imag

Imagination Augmented

- … uses rollouts to imagine the best fu
- The IAP consists of **one Imaginatior**
- All Imagination Rollout outputs C_i wil

Single Rollout

- … evaluates **how a selected action**
- Imagines the future by chaining multi the current state as well as a start ac the next Imagination Core.
- After performing n rollout steps a **cor** Imagination Rollout

Imagination Core (IC)

Imagination Core (IC)

- … predicts the next state based on a
- Consists of a Policy Net and a pretra
- The policy net predicts the next actio roposed by [1] we used A3C [2] as ou
- Output: predicted reward $\hat{r}_{t+1}$ and th

Model Free Path

- … provides the network with an optio **deal with insufficient future predic**
- Uses the **convolutional layers of A3** model free architecture [2] but does n include the fully connected layer

Literature

[1] Racanière, Sébastien, et al. "Imagination-Augmented Agents for D
[2] Mnih, Volodymyr, et al. "Asynchronous methods for deep reinforce
[3] Leibfried, Felix, Nate Kushman, and Katja Hofmann. "A deep learn arXiv:1611.07078 (2016).
[4] Brockman, Greg, et al. "Openai gym." arXiv preprint arXiv:1606.01
[5] https://github.com/mpSchrader/I2A-for-Deep-RL

Augmented Ag
ent Learning

hober, Florian Klemt, Max-Philip

magination Augment

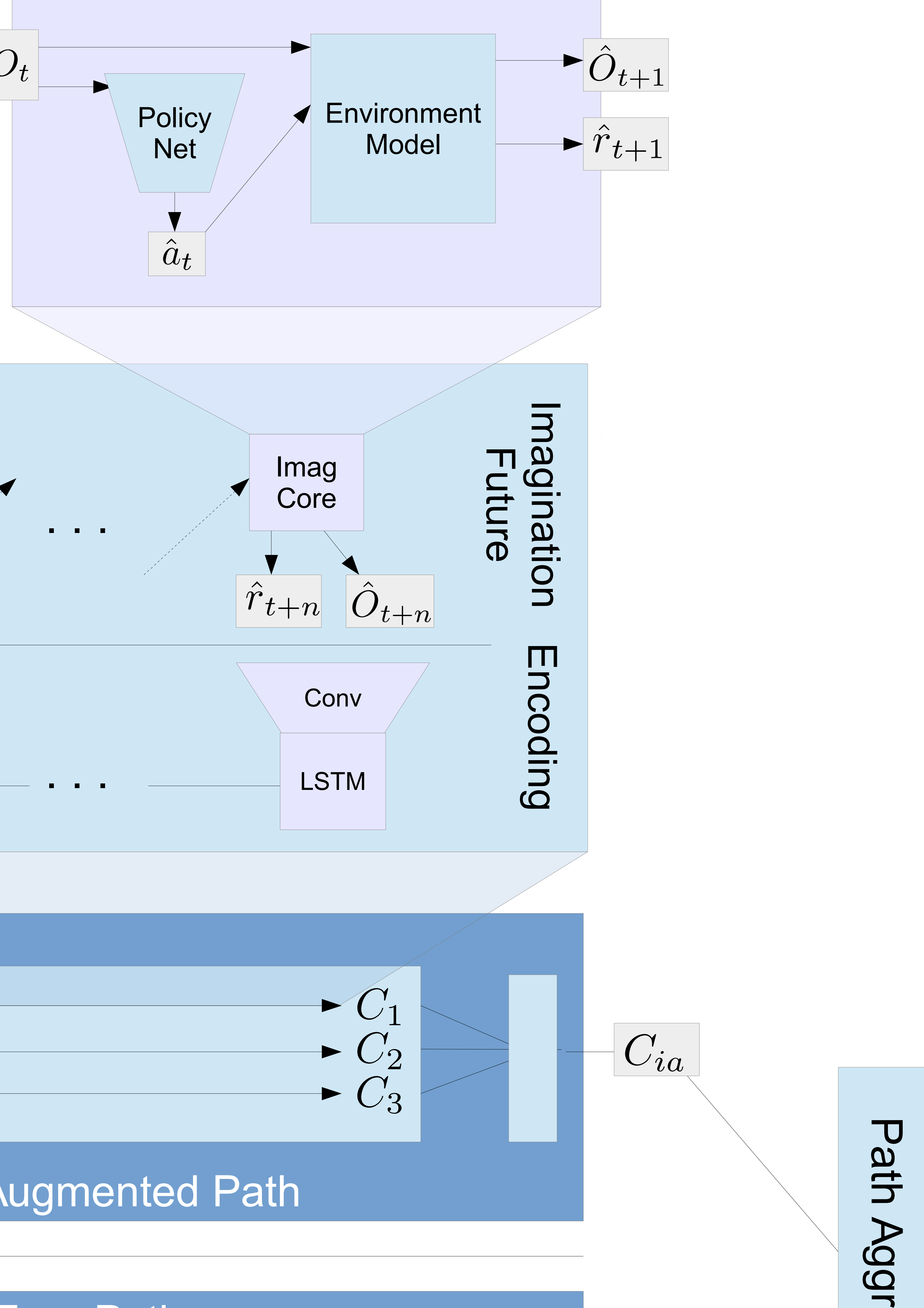
r *Imagination Augmented Agents for Deep*

ts of DeepMind using their proposed desig
Atari games and we used **OpenAI Gym a**

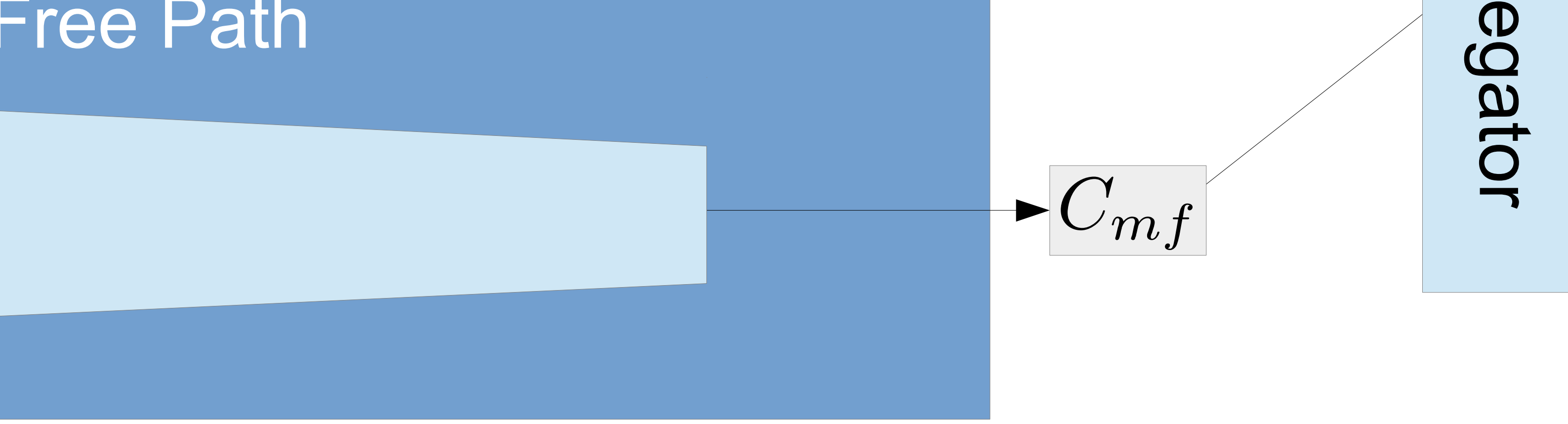
ree Reinforcement Learning Architectures

re an imagined future of available actions

Imagination Core



$\hat{O}_{t+1}$
$\hat{r}_{t+1}$
Policy Net
Environment Model
$\hat{a}_t$
Imag Core
$\hat{r}_{t+n}$
$\hat{O}_{t+n}$
Imagination Future
Conv
LSTM
Encoding
C_1
C_2
C_3
C_{ia}
Path Aggr



gination Augmented Agent – Architecture

Path (IAP)

uture action

n **Rollout** for all available **actions** a_i

l be aggregated by concatenating them to

performs in the future

ple imagination cores. At the beginning it ta
tion. Finally the predicted state $\hat{O}_{t+1}$ gets

nvolutional LSTM encodes the result of n internal selected action $\hat{a}_t$

ined Environment Model

n to perform, the policy net is a simple poli
ur policy net

e next state $\hat{O}_{t+1}$

on to
tions

3C
not

Path Aggregator

- … calculates based on both p
  π and value V
- First, the output of the paths $($
  gets concatenated
- This then is followed by a fully
  net which outputs the policy a

eep Reinforcement Learning." Advances in Neural Information Processing S
ment learning." International Conference on Machine Learning. 2016.
ing approach for joint video frame and reward prediction in atari games." ar
540 (2016).

ents for Deep

pp Schrader

ed Agent Architecture

Reinforcement

gn choices, as
s an Atari

s

Environment M

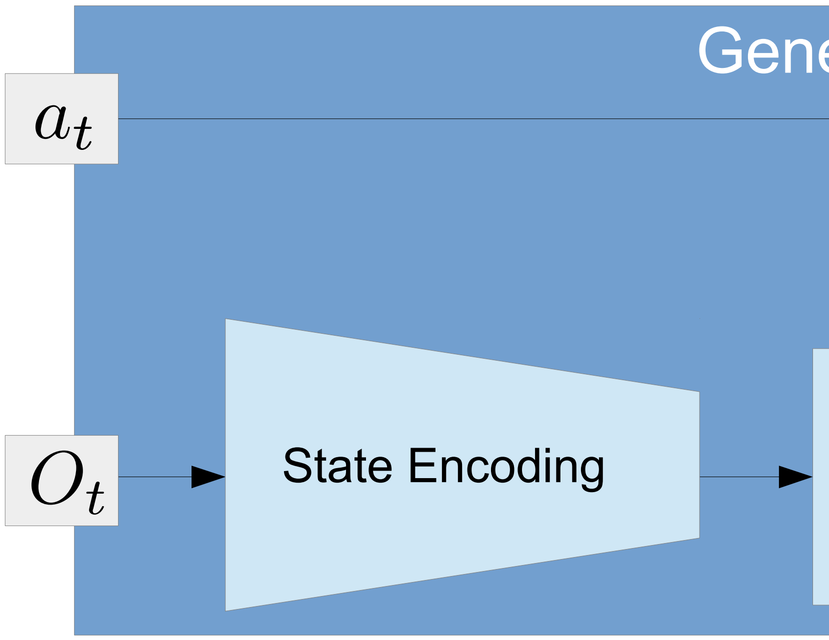


Fig. 2:

- **predicts the nex**

… predicts the nex

- The Environment Mo
in the paper due to di

- We used the architec
takes one hot encode
as input and predicts

- In the latent space th
are combine by elem

- For training we found
in combination with **A**
generate the best res

Input

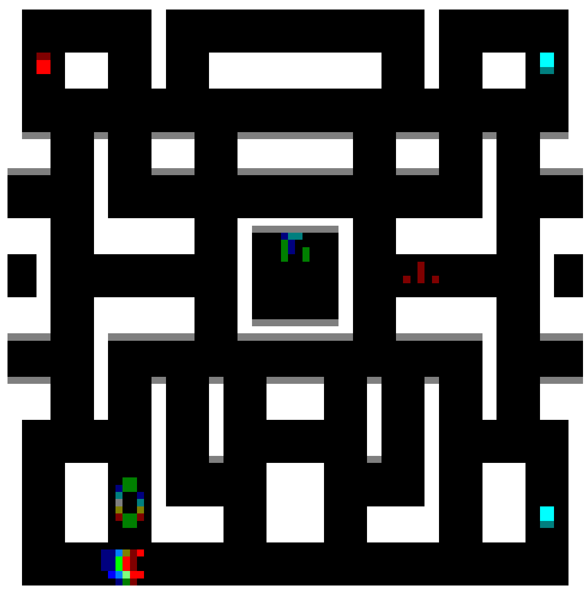
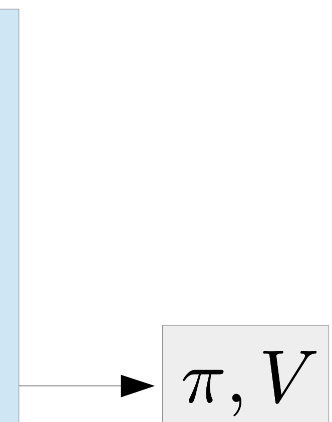


Fig. 3: Predict



$$C_{im}$$

akes
passed into

the

E

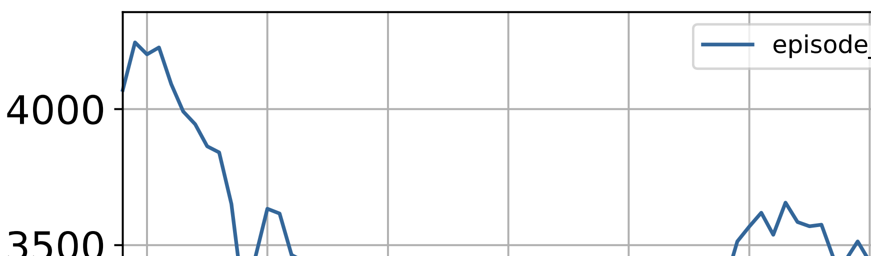
- **Scott Reed**, DeepMin
ambitions project” ar
PacMan?”

- For **training** the I2A n
method proposed in D

- Due to computational
very strong model. De
Atari environment step

- Never the less we wer
Pong, but we can **not**
Architecture **with suc**

- We saw a changing du
learning and it’s ability



cy net. As

paths a policy

C_{im} and C_{mf}

y connected
and the value

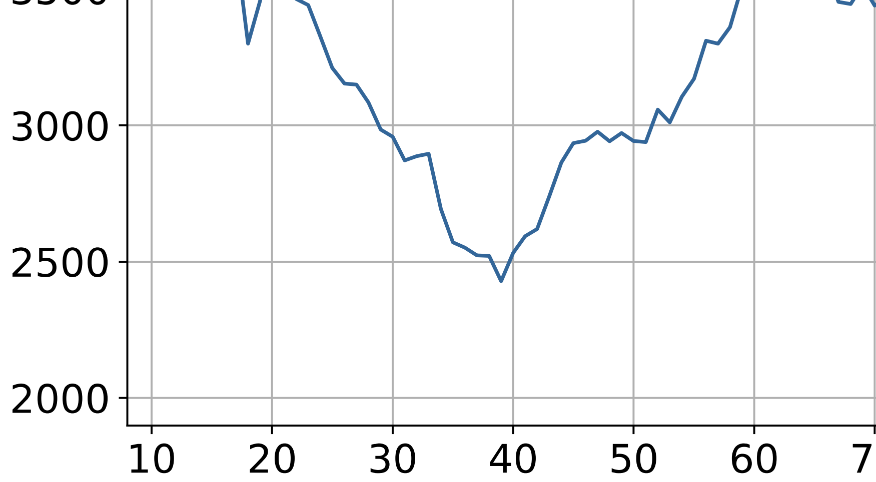


Fig. 4: Episodes played per game c

S

- As described above, w
strong model, but we w
model, which is able t

- Our code will be publis
after the class

Systems. 2017.

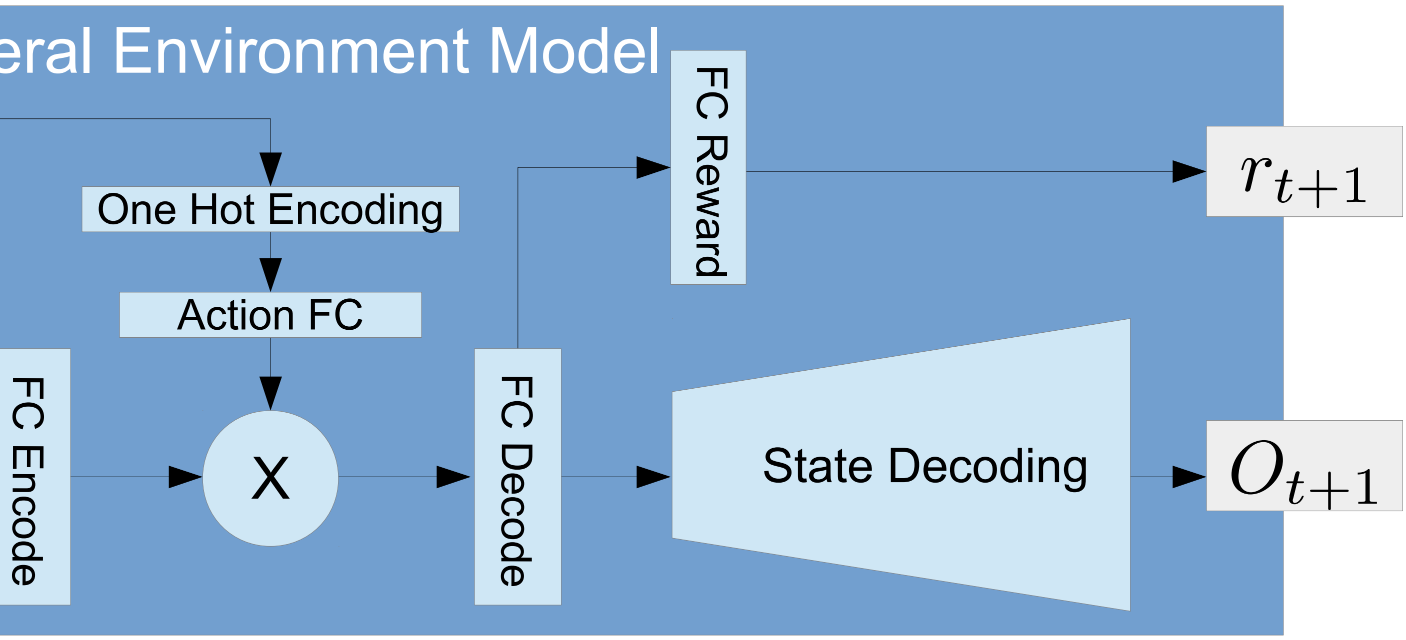
Xiv preprint

Get in Touch

Angela Denninger	angela.dennin
Felix Schober	felix.schober(
Florian Klemt	florian.klemt@
Max-Philipp Schrader	mschr@mit.e

e

Model



Architecture of the Environment Model

t state and reward

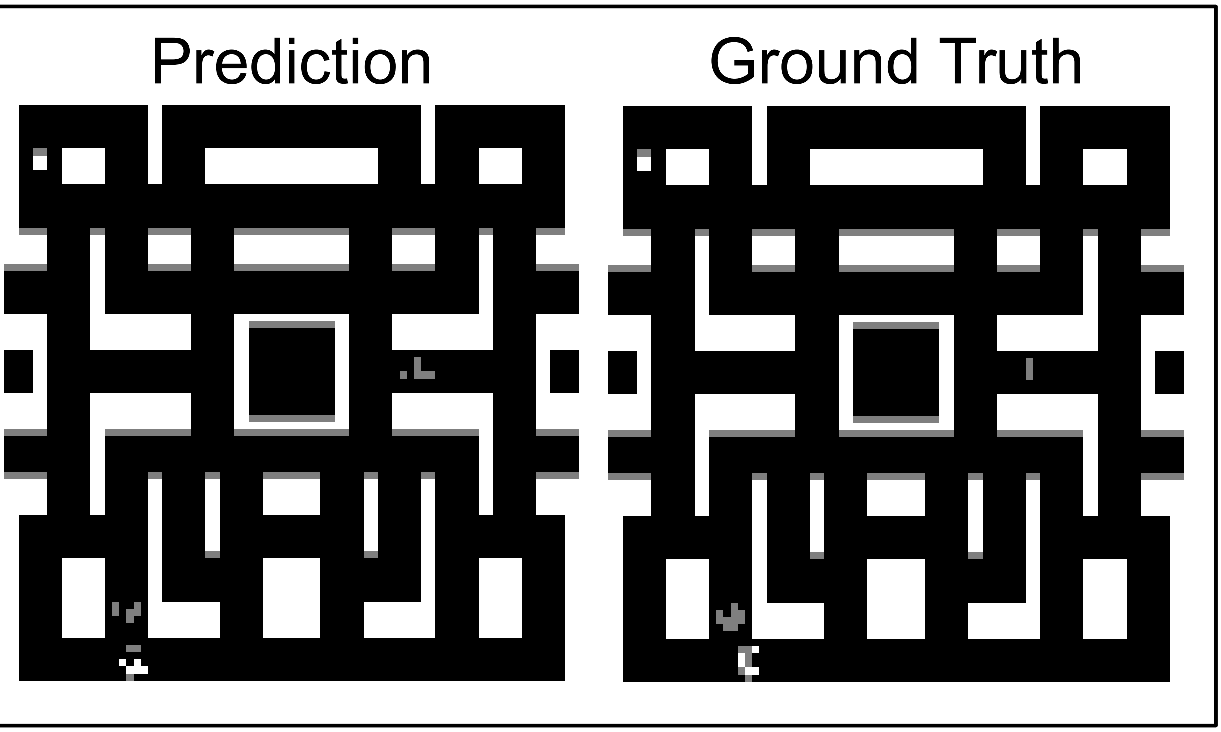
t state and reward

del differs from the ones proposed
ifferent environment state sizes

cture proposed in [3]. The model
ed actions and the current frame
next state and reward

e Action FC and the State Encoding
ent wise multiplication

l the **Negative Log Likelihood Loss**
Adam and a **learning rate of 10^{-4}**
sults



ion Example for the MsPacman Environment

Evaluation

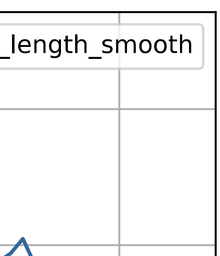
nd, 01/30/2017: “Oh… **That’s a very** nd “What you want to use real

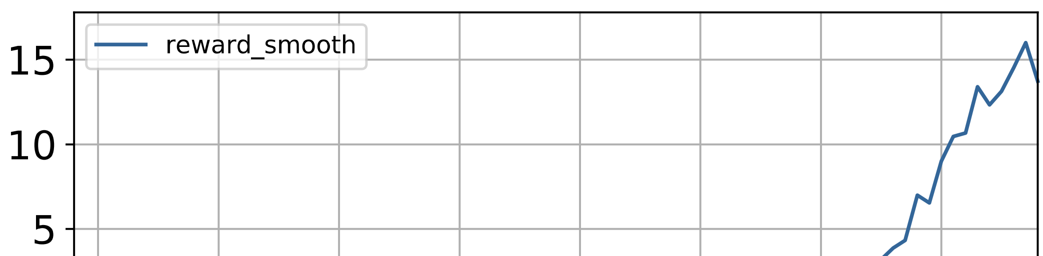
etwork we used the **asynchronous** DeepMinds A3C paper [2]

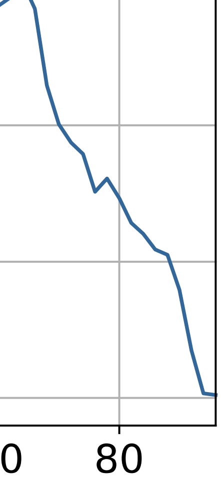
resources, we were not able to train a eepMind trained their I2A model for 10^9 os.

re able to **train a working version for** t **proof the advantages** of the I2A- **h a simple model**

uration length, which is based on y to win faster in the end.







of Pong

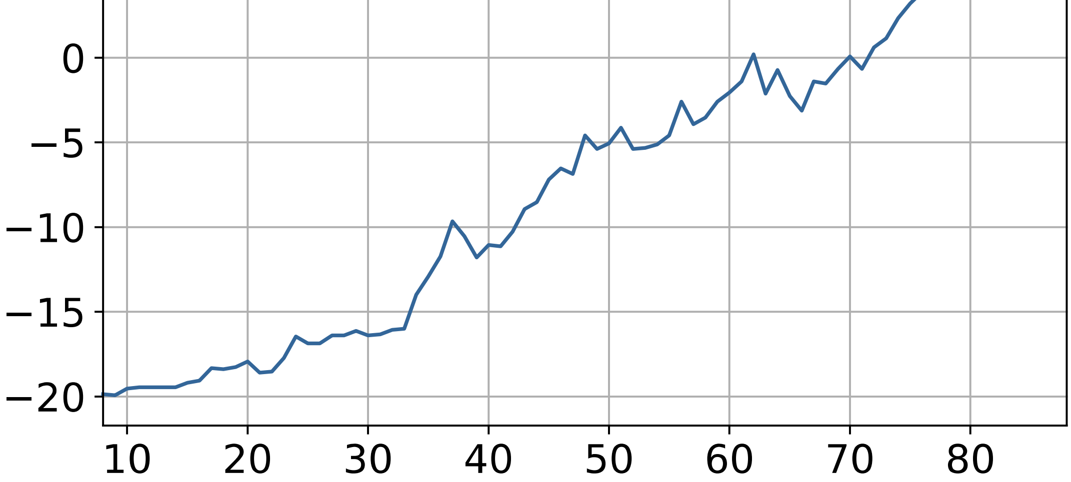


Fig. 5: Reward archived per game of Pong

Summary

we were not able to train a sufficiently
were able to **implement a working I2A**
o **learn and play Atari Games**

shed as **Open-Source on Github** [5]

nger@tum.de
@tum.de
@tum.de
du