Master Thesis Kolloquium

A new Adversarial Approach for Model Generation in both a Supervised and Reinforcement Learning Context

Christian Coenen

Agenda

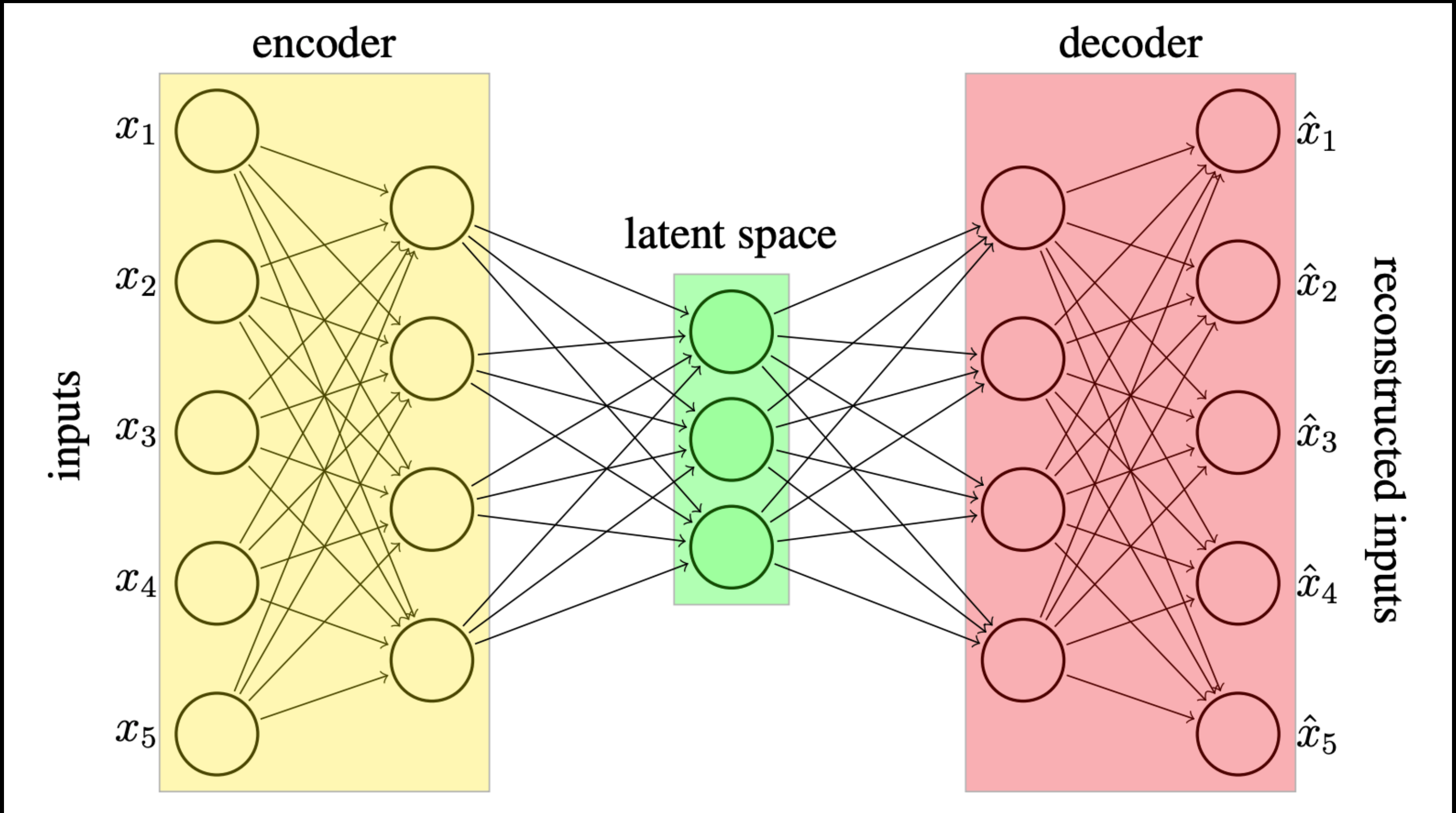
- Motivation
- Objectives
- Fundamentals
- Objective 1
    - Concept -> Results -> Conclusion -> Outlook
- Objective 2
    - Concept -> Results -> Conclusion -> Outlook

Motivation

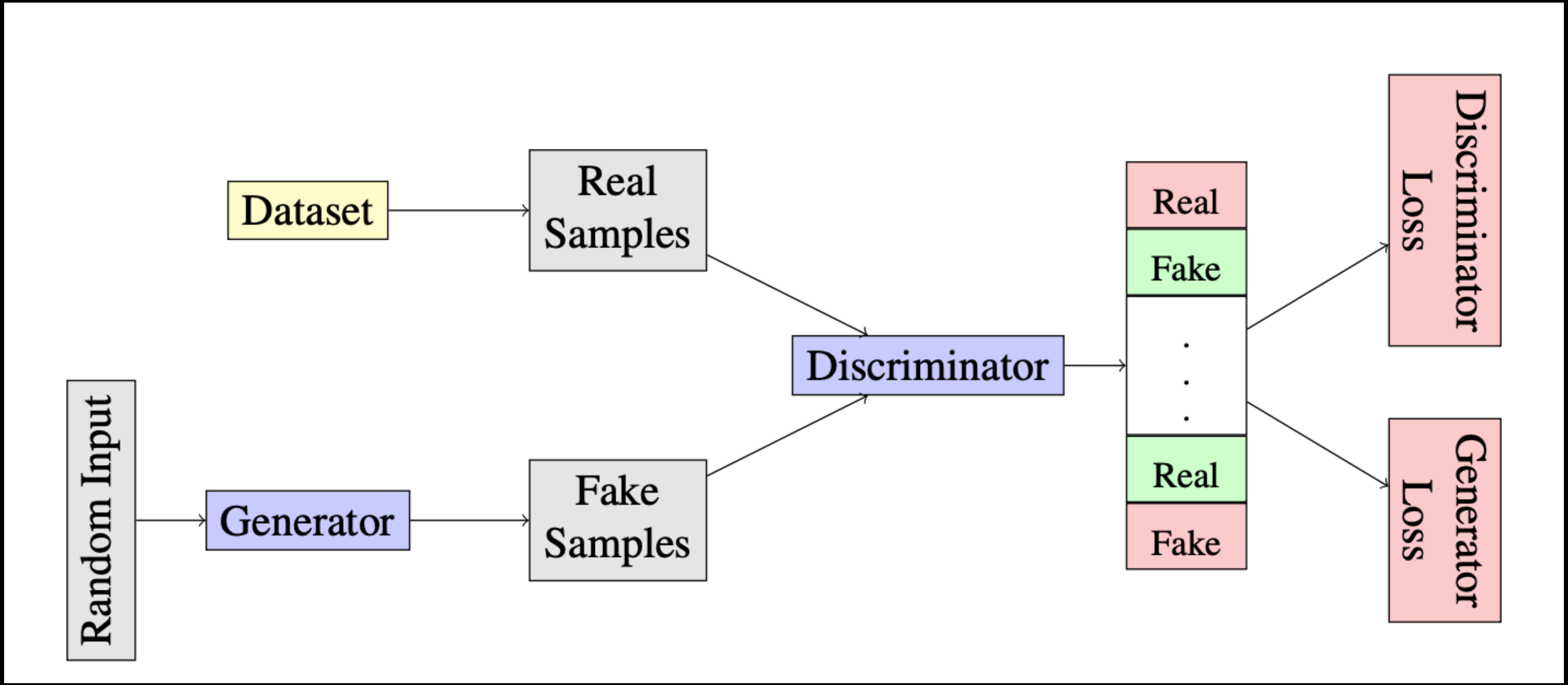
https://worldmodels.github.io/

Fundamentals

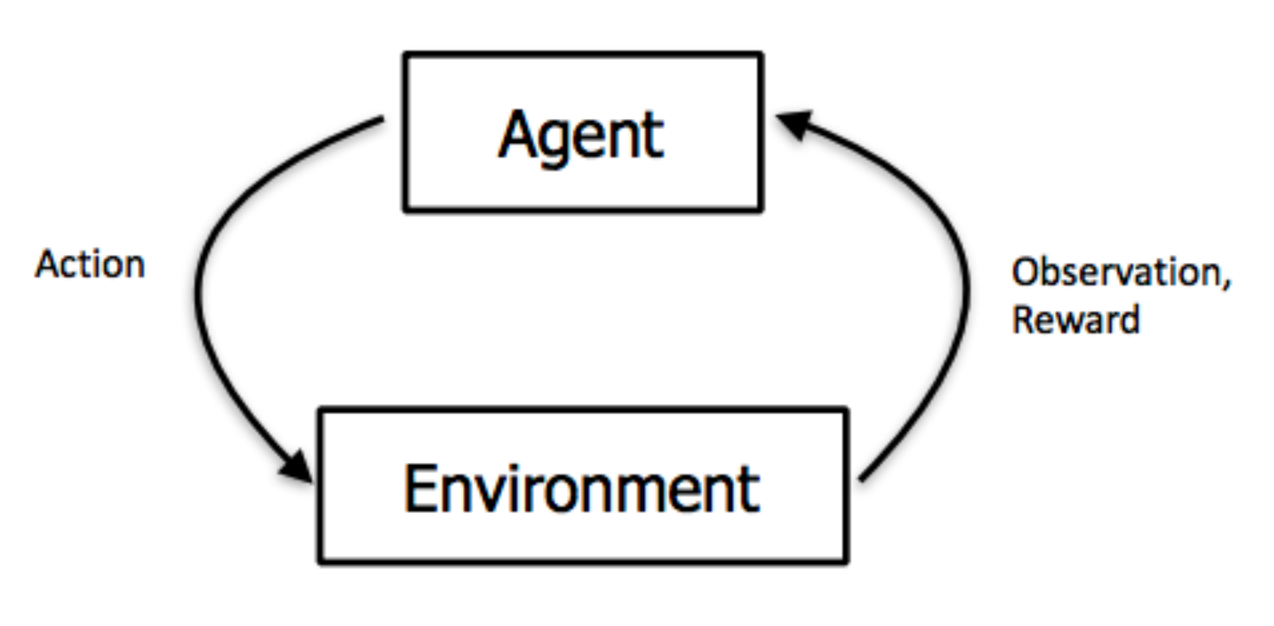
Autoencoder (AE)



Generative Adversarial Network (GAN)



Reinforcement Learning



Objectives (O)

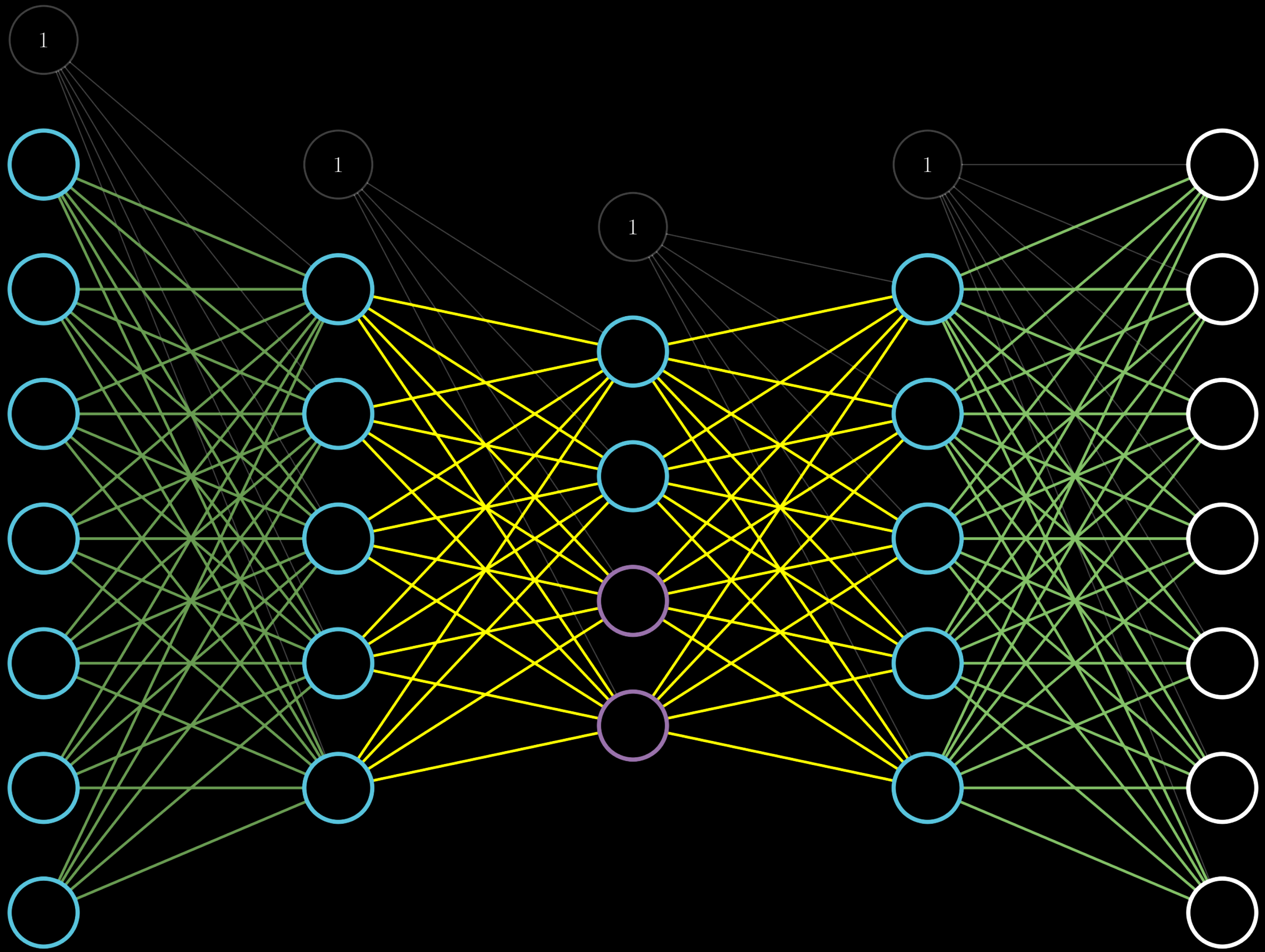
- O1: Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights.

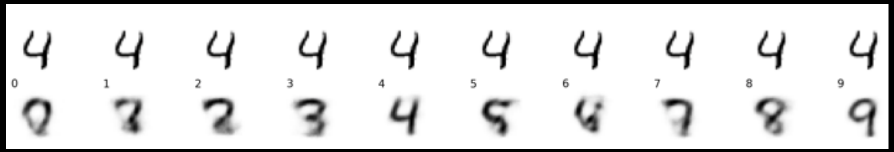
- O2: Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment.

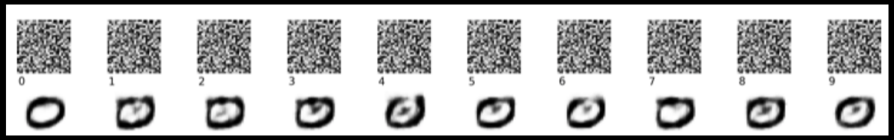
Concept O1

*”Generating new samples given random inputs and a label,
as well as encoding these samples, using the same weights.”*

Classification Autoencoder (CAE)

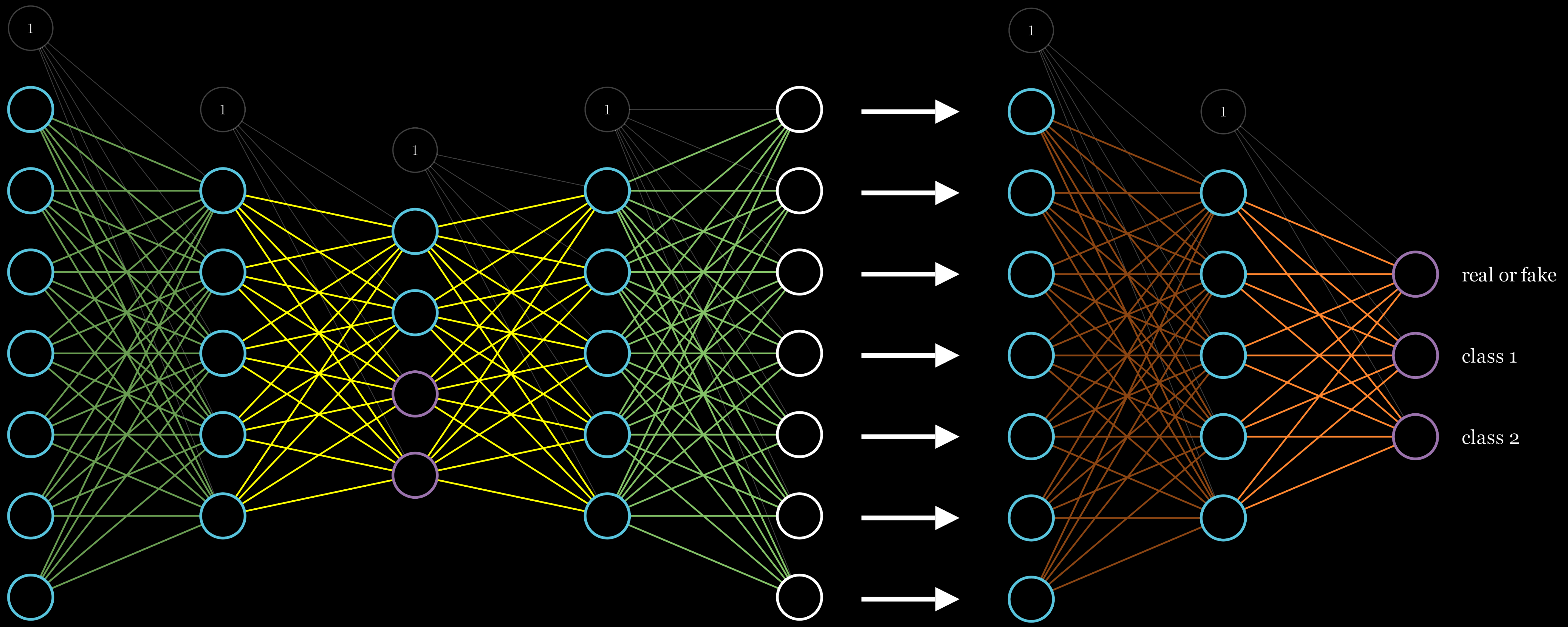






"Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights."

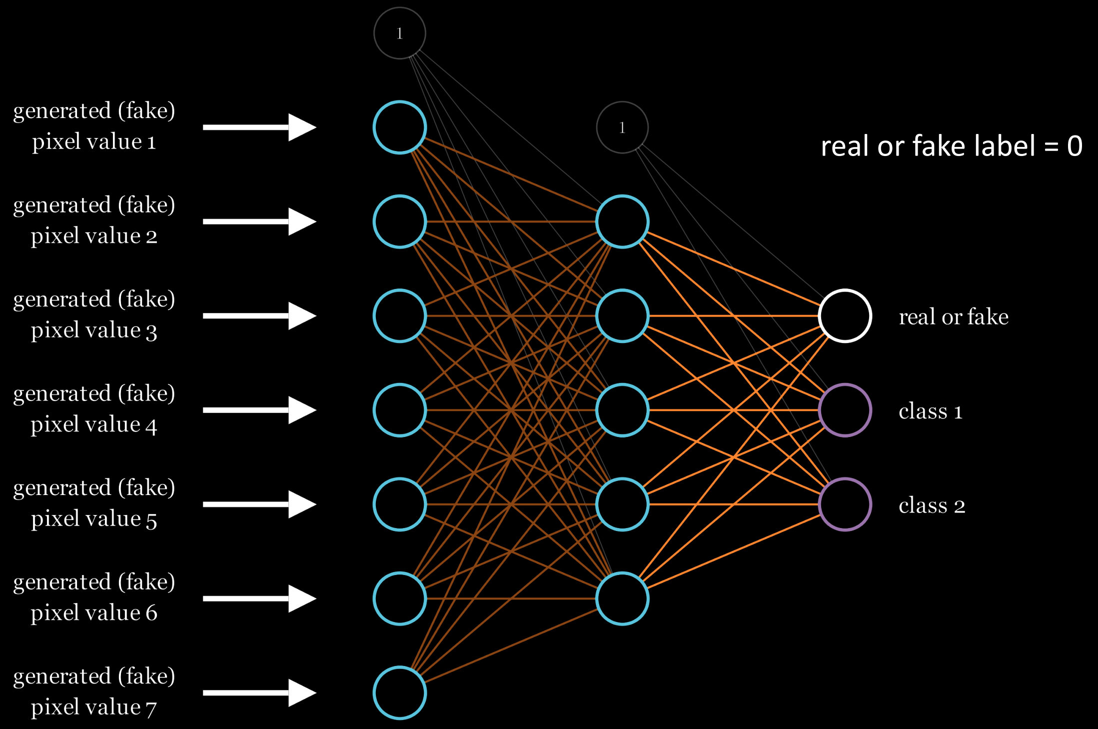
Shared Aversarial CAE (SA-CAE)



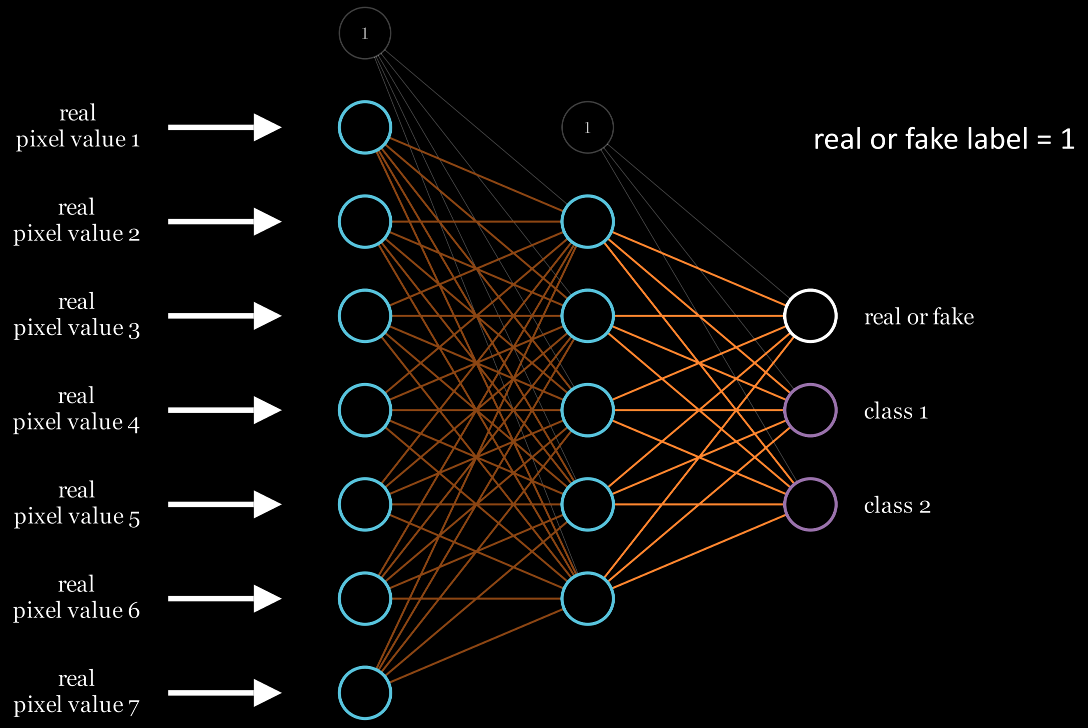
"Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights."

Training Step 1 – Discriminator

batch / 2

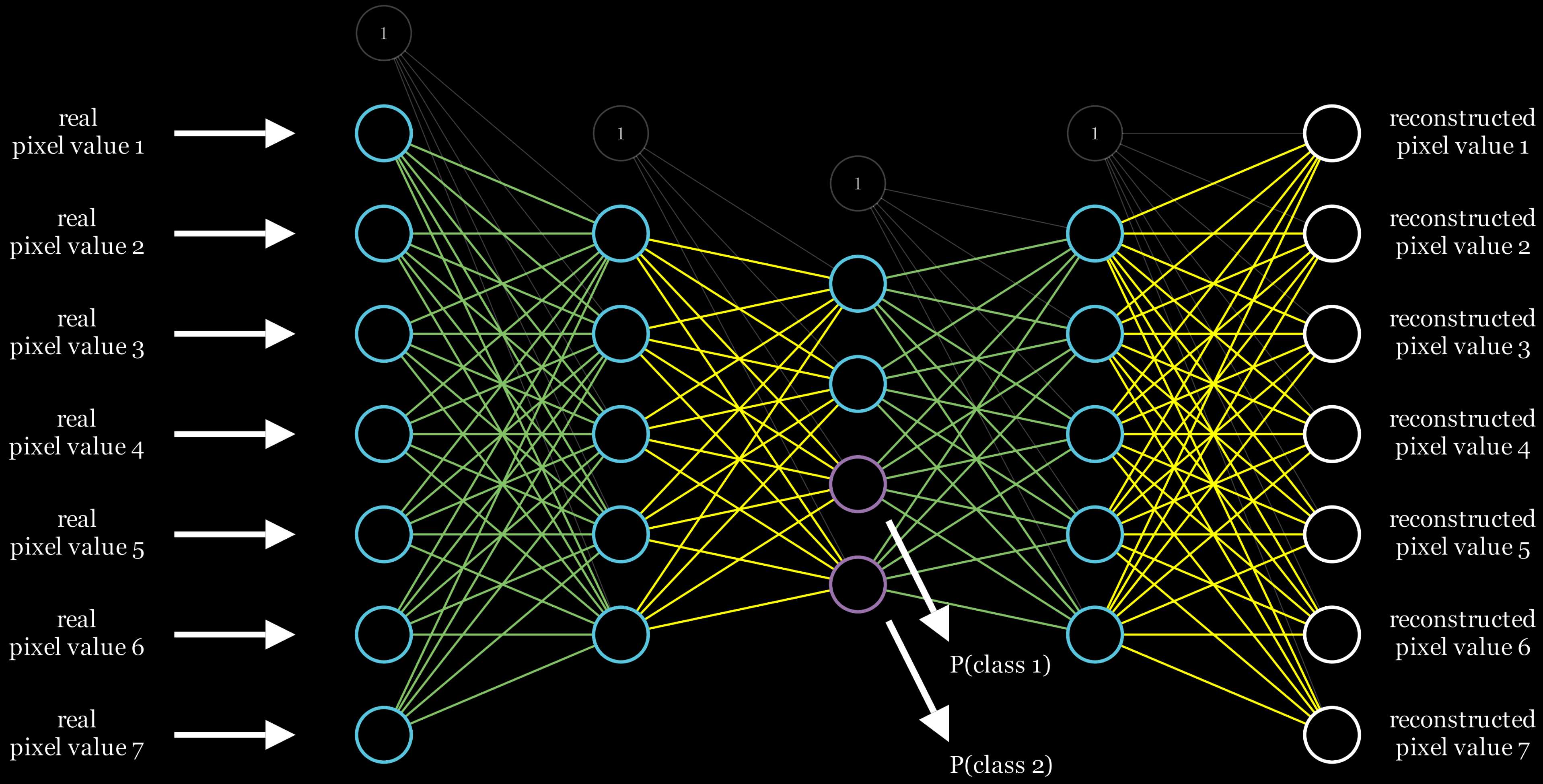


batch / 2



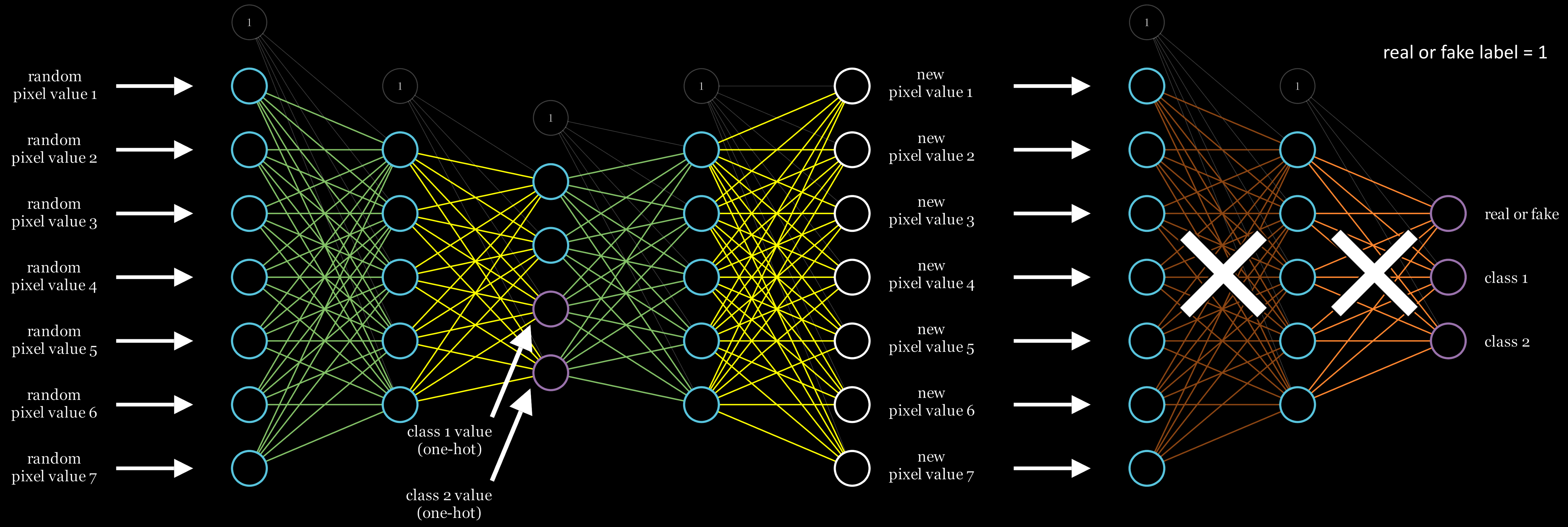
"Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights."

Training Step 2 – Generator Supervised



"Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights."

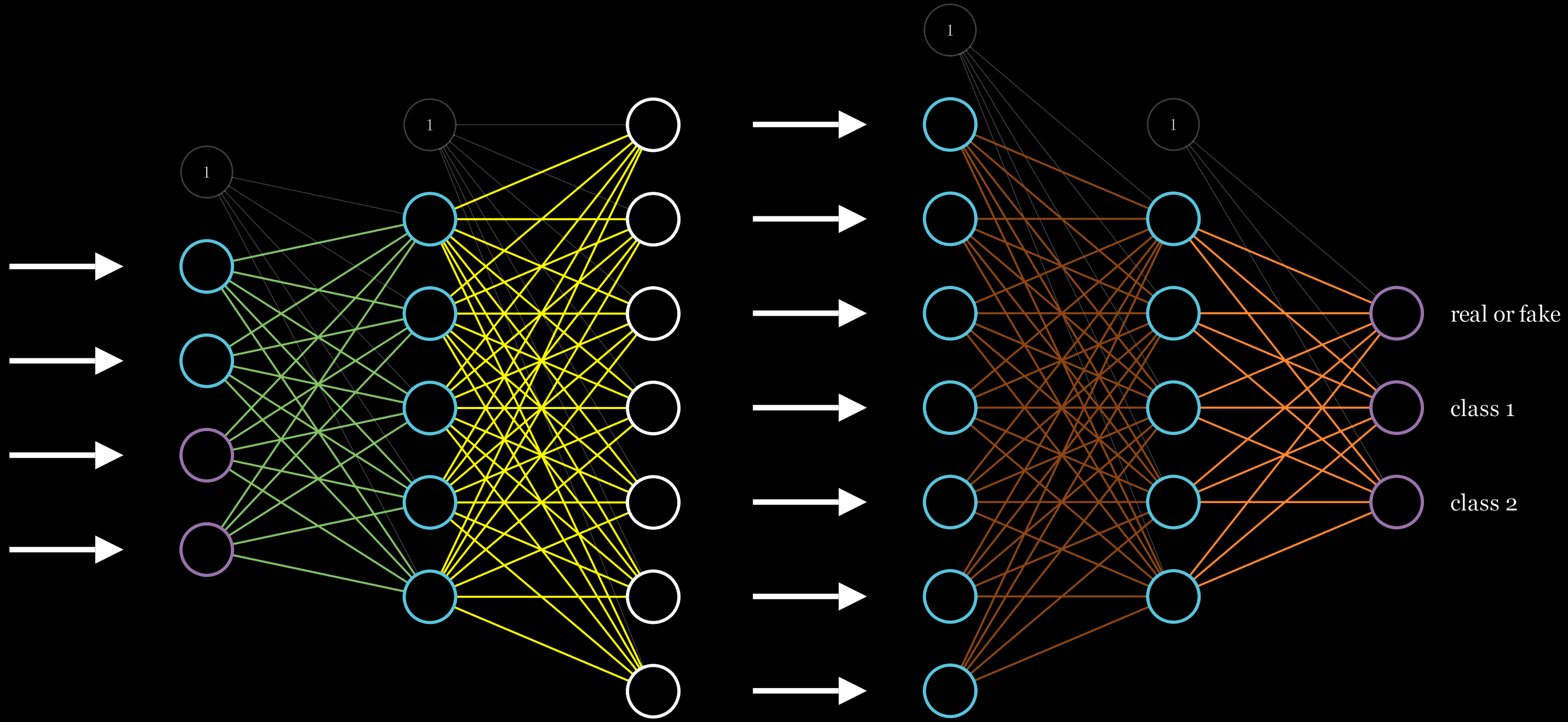
Training Step 3 – Generator via GAN



= not updated during backpropagation

"Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights."

AC-GAN (used for comparison)



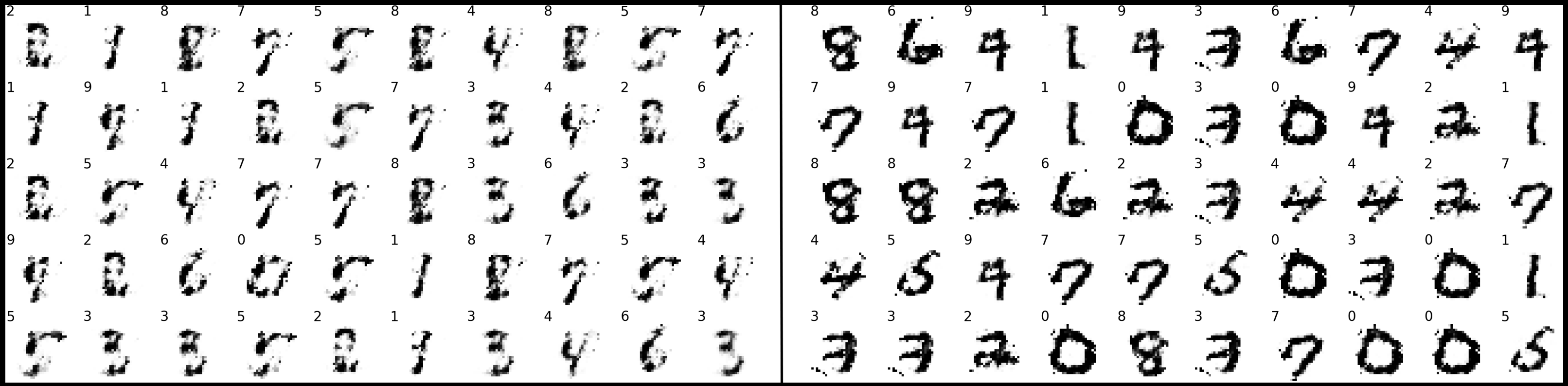
"Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights."

Results

AC-GAN

6 epochs

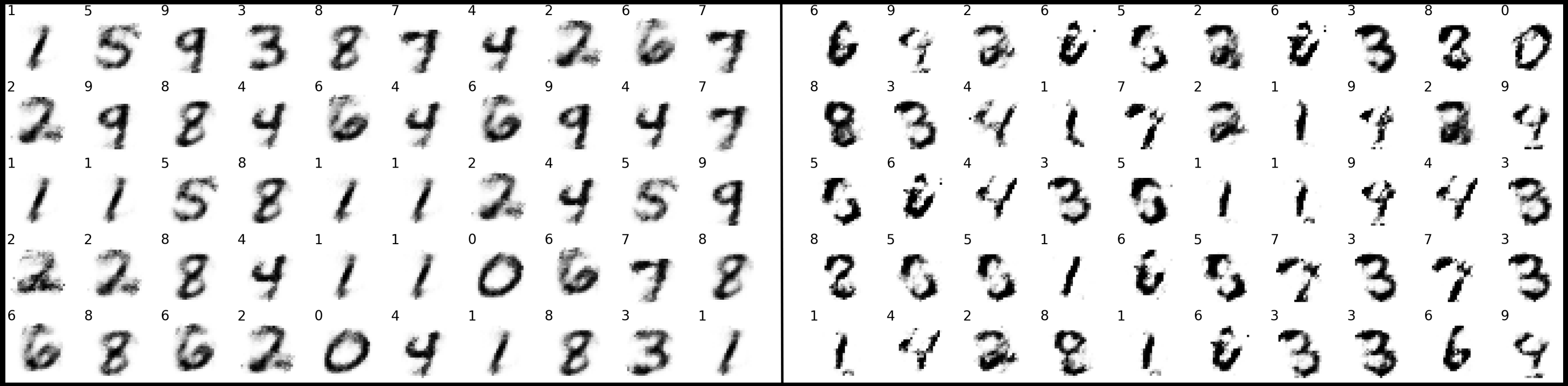
30 epochs



SA-CAE

6 epochs

30 epochs



”Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights.”

Conclusion

SA-CAE	AC-GAN
unstable	stable
faster learning	slower learning
addresses the objective	does not address the objective (no encoding, weight sharing)
usable for image synthesis (completion, correction)	not usable for image synthesis

”Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights.”

Outlook

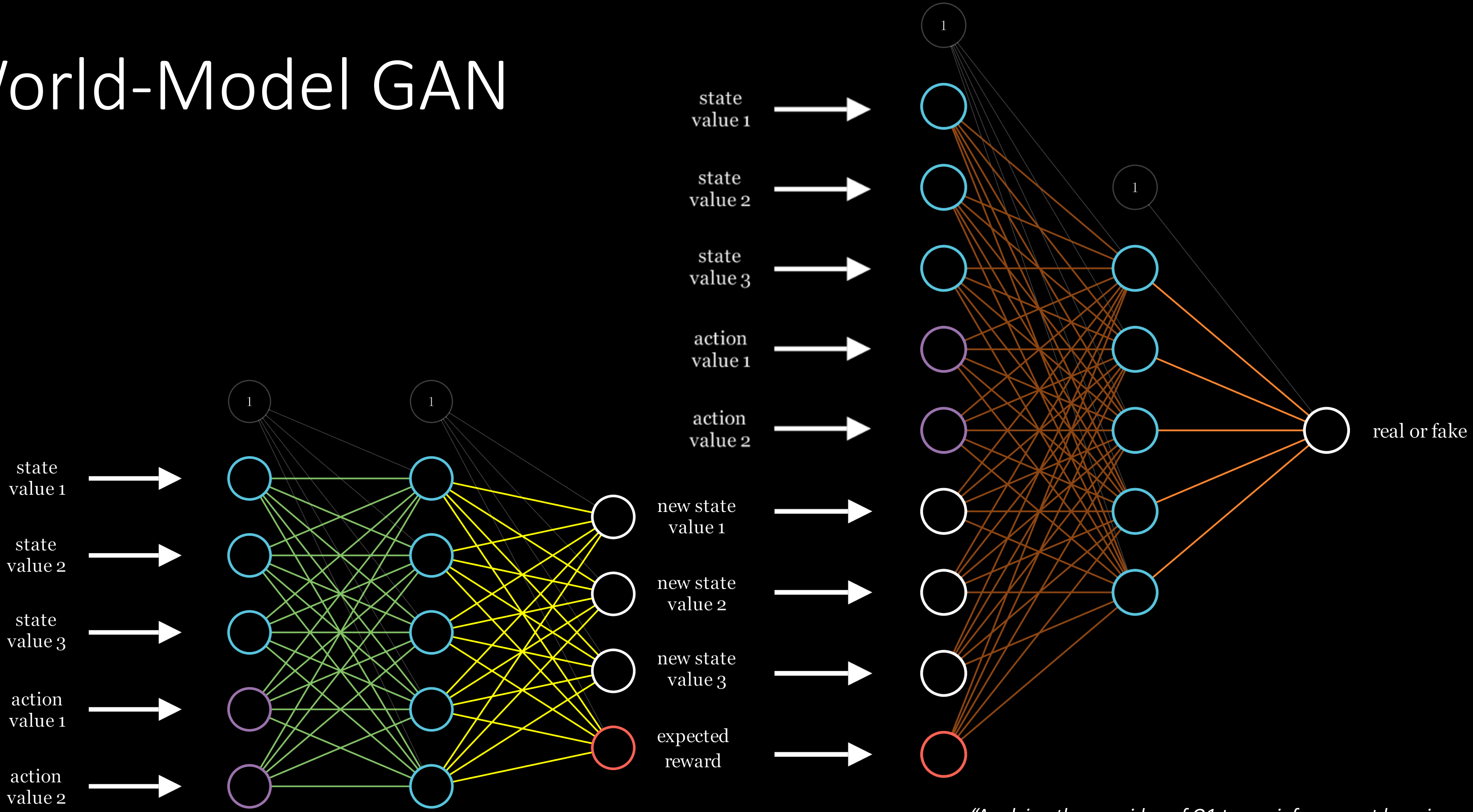
- Convolutional layers
  - MS-SSIM (requires other dataset)
  - allows comparison with other approaches
- Shared classifier for Autoencoder and GAN
  - might result in an improved classifier due to newly generated samples during discriminator training

”Generating new samples given random inputs and a label, as well as encoding these samples, using the same weights.”

Concept O2

“Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment.”

World-Model GAN



"Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment."

Training approaches

train data

test data

Approach 1

1. Discriminator training
    - Real data
    - Fake data (generator)
2. Generator supervised
3. Generator via GAN

Approach 2

1. Discriminator training
    - Real data
    - Fake data (generator)
2. Generator supervised
3. Generator via GAN

Approach 3

1. Discriminator training
    - Real data
    - Fake data (generator)

~~2. Generator supervised~~

3. Generator via GAN

"Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment."

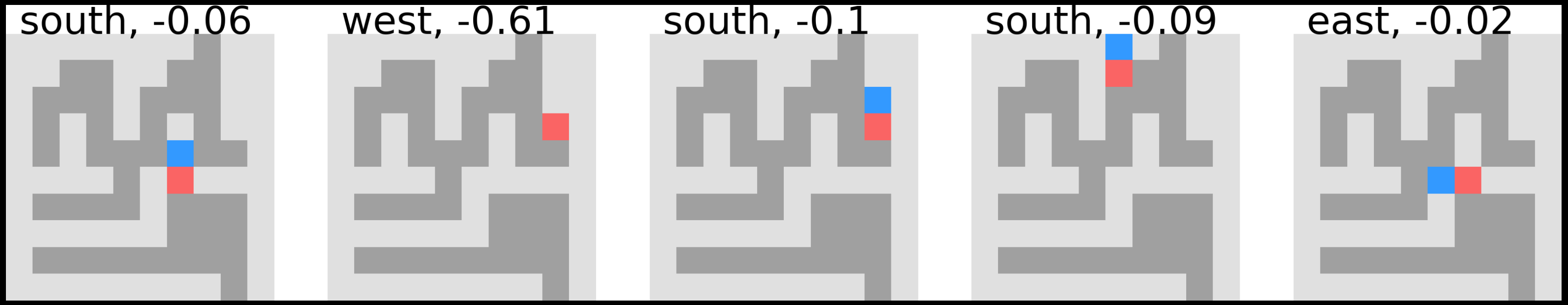
Results

= input state

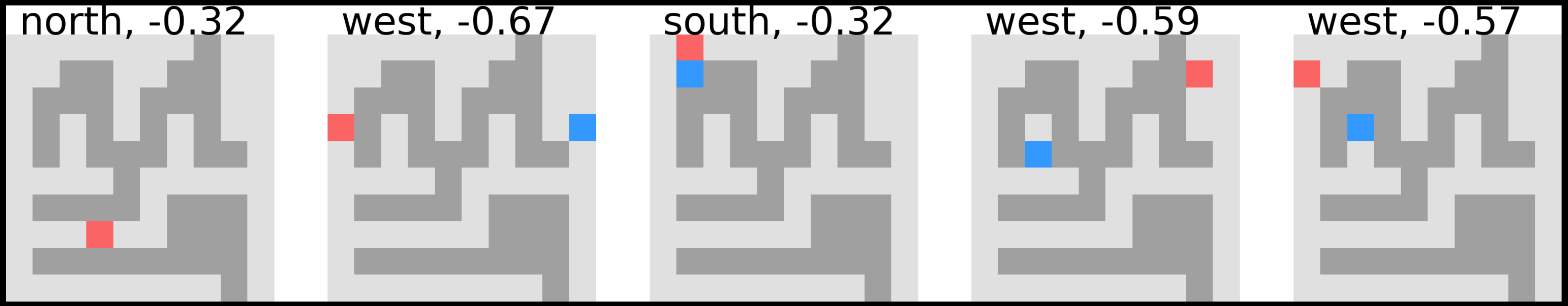
= predicted next state

overlaps

Train data inputs



Test data inputs



Approach 1

1. Discriminator training
   - Real data
   - Fake data (generator)
2. Generator supervised
3. Generator via GAN

"Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment."

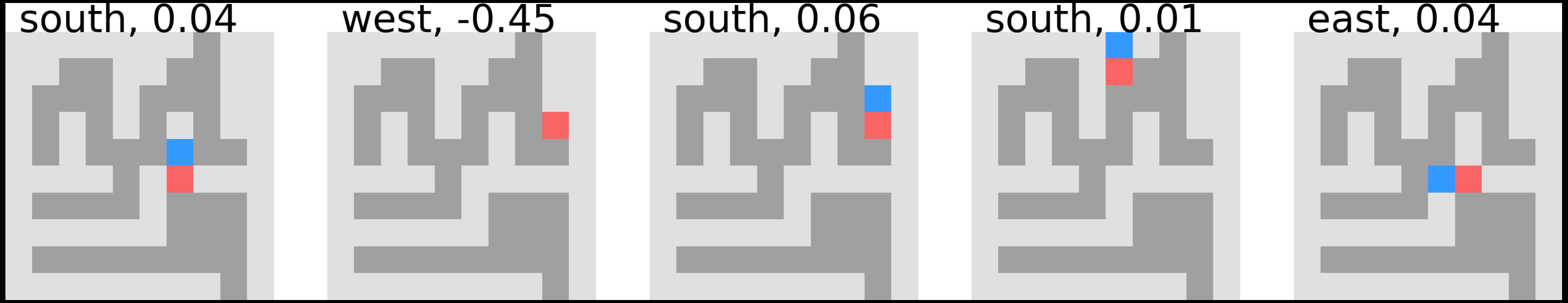
Results

= input state

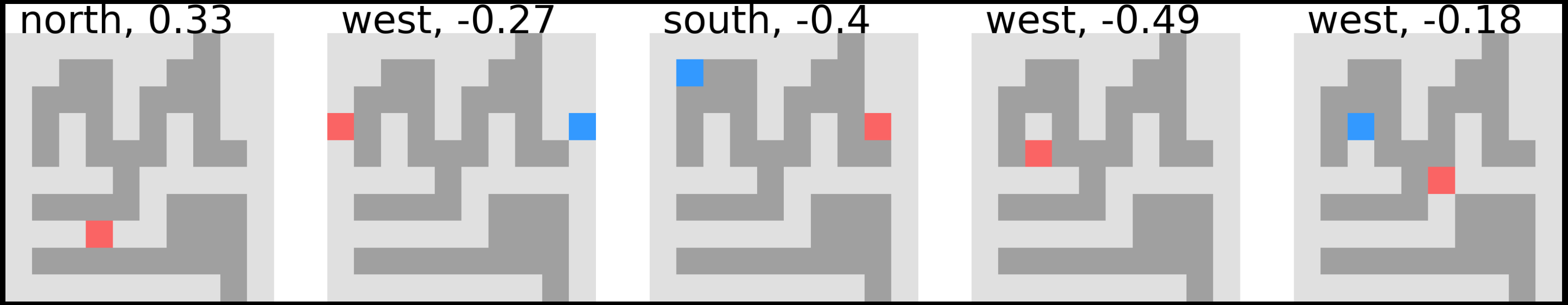
= predicted next state

overlaps

Train data inputs



Test data inputs



Approach 2

1. Discriminator training
   - Real data
   - Fake data (generator)
2. Generator supervised
3. Generator via GAN

"Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment."

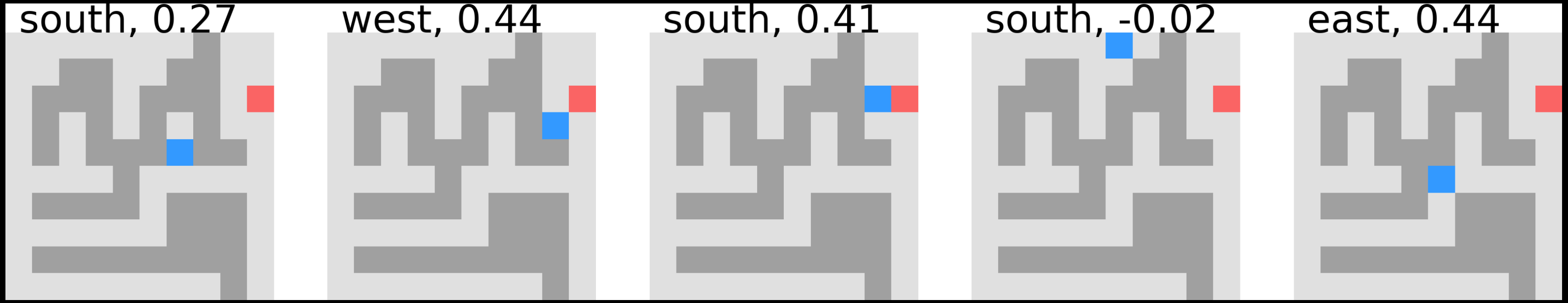
Results

= input state

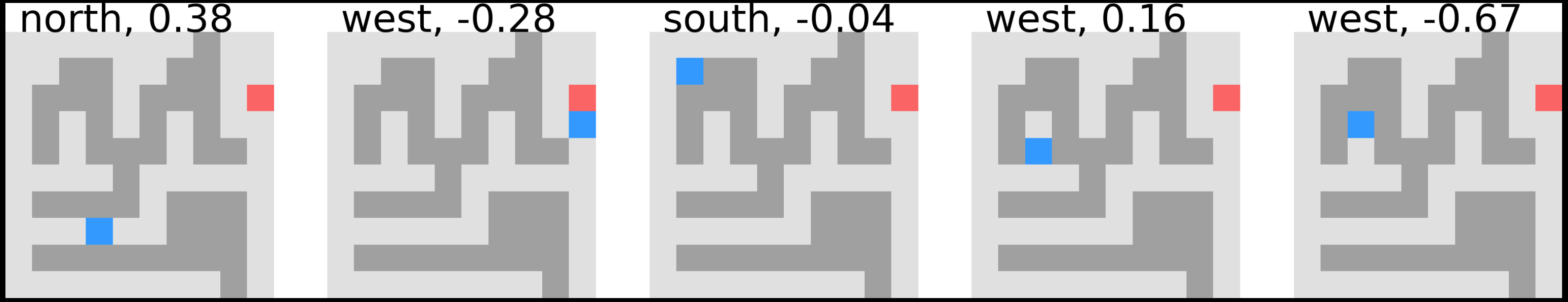
= predicted next state

overlaps

Train data inputs



Test data inputs



Approach 3

1. Discriminator training
   - Real data
   - Fake data (generator)

~~2. Generator supervised~~

3. Generator via GAN

"Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment."

Conclusion

- Accurate world model can be learned for experienced state-action pairs

- Unable to learn the physics of the environment
  - inacurate predictions for unseen state-action pairs

=> further research required

"Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment."

Outlook

Assuming the physics of the environment can be learned:

- cyclic training process between sampling the environment and model improvement by going through new trajectories
- Q-value prediction can be integrated into the generator network

=> Agent can sample new trajectories using its own world model

"Applying the core idea of O1 to a reinforcement learning setup to generate trajectories of realistic new states, not present in the dataset, without using the environment."

Thank you for your attention!