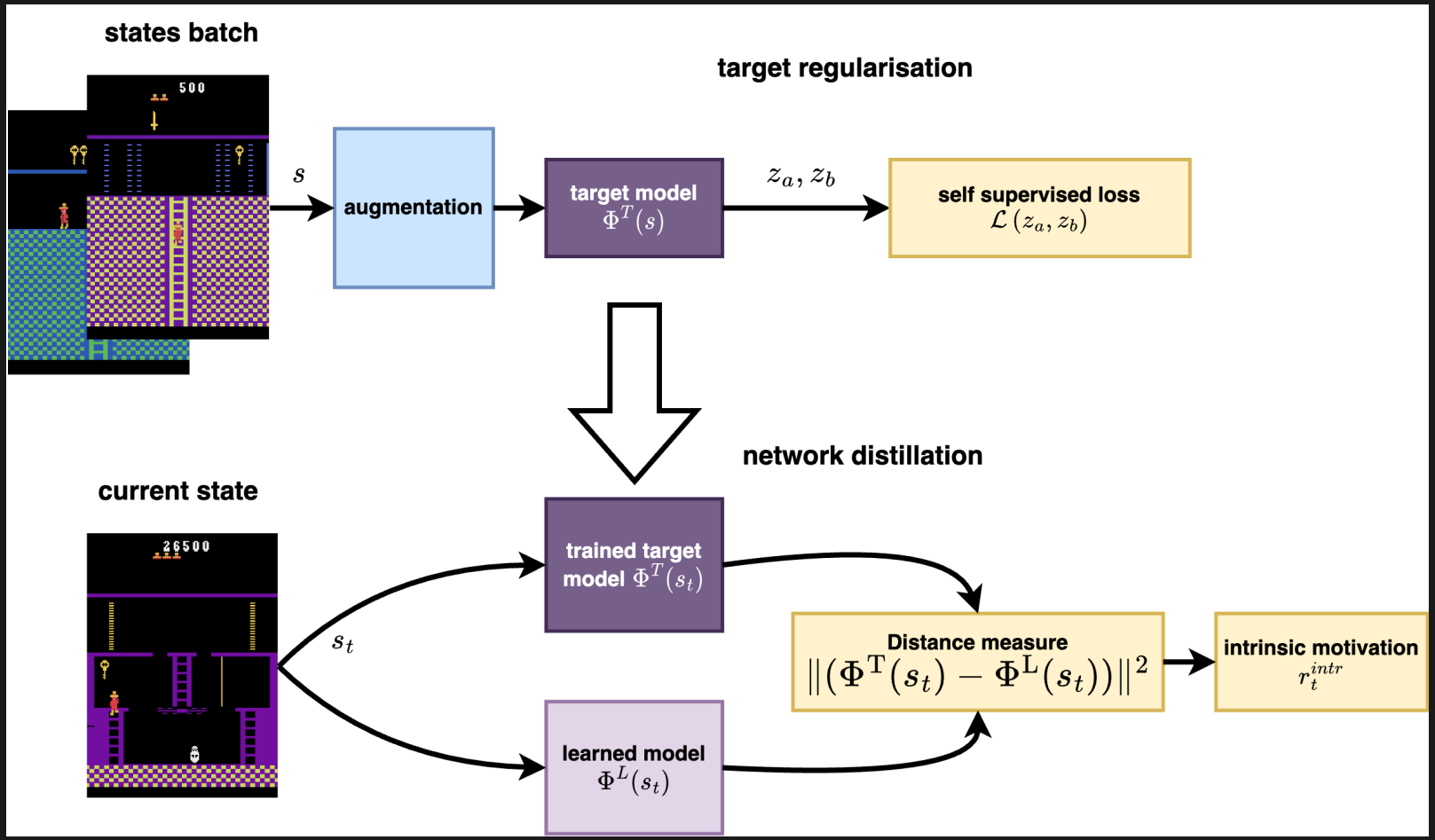
Exploration by self supervised-exploitation

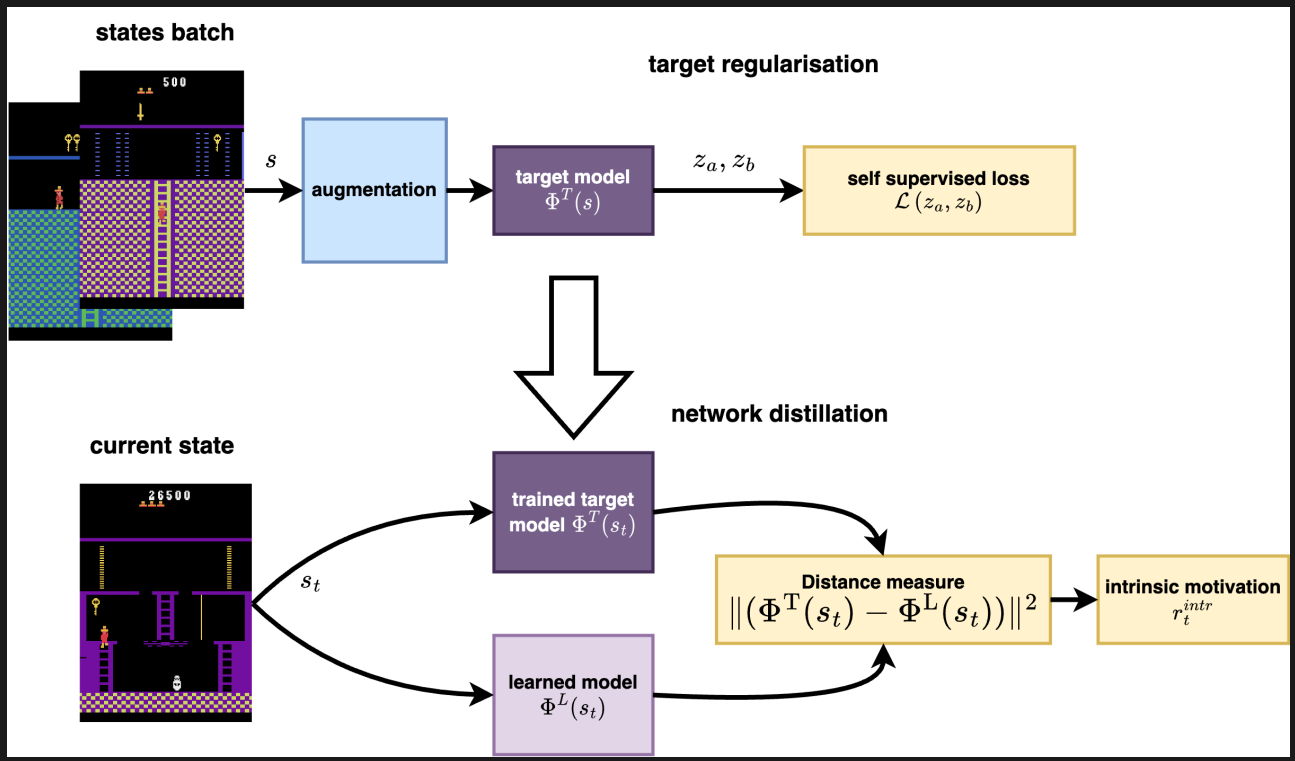
Matej Pecháč
Michal Chovanec
Igor Farkaš

Department of Applied Informatics
Comenius University in Bratislava
Slovak Republic

Exploration by self supervised-exploitation

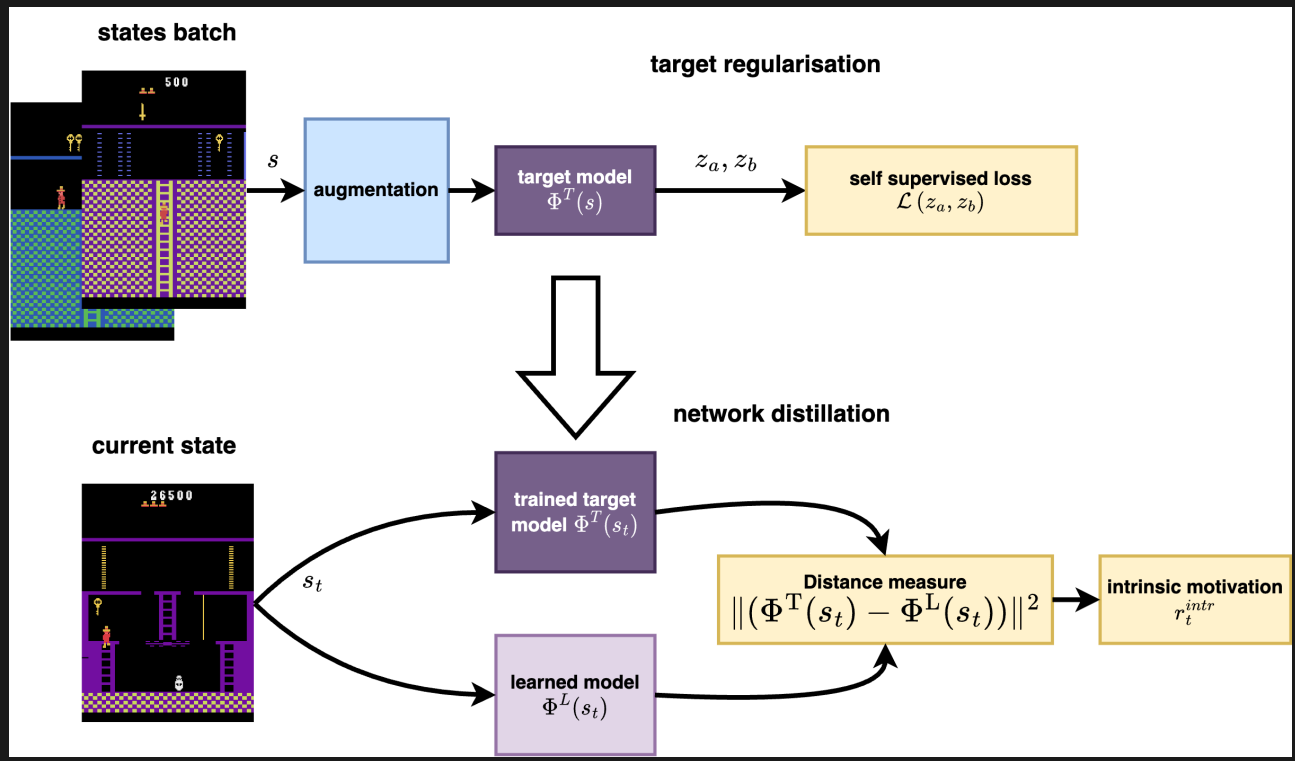


Exploration by self supervised-exploitation



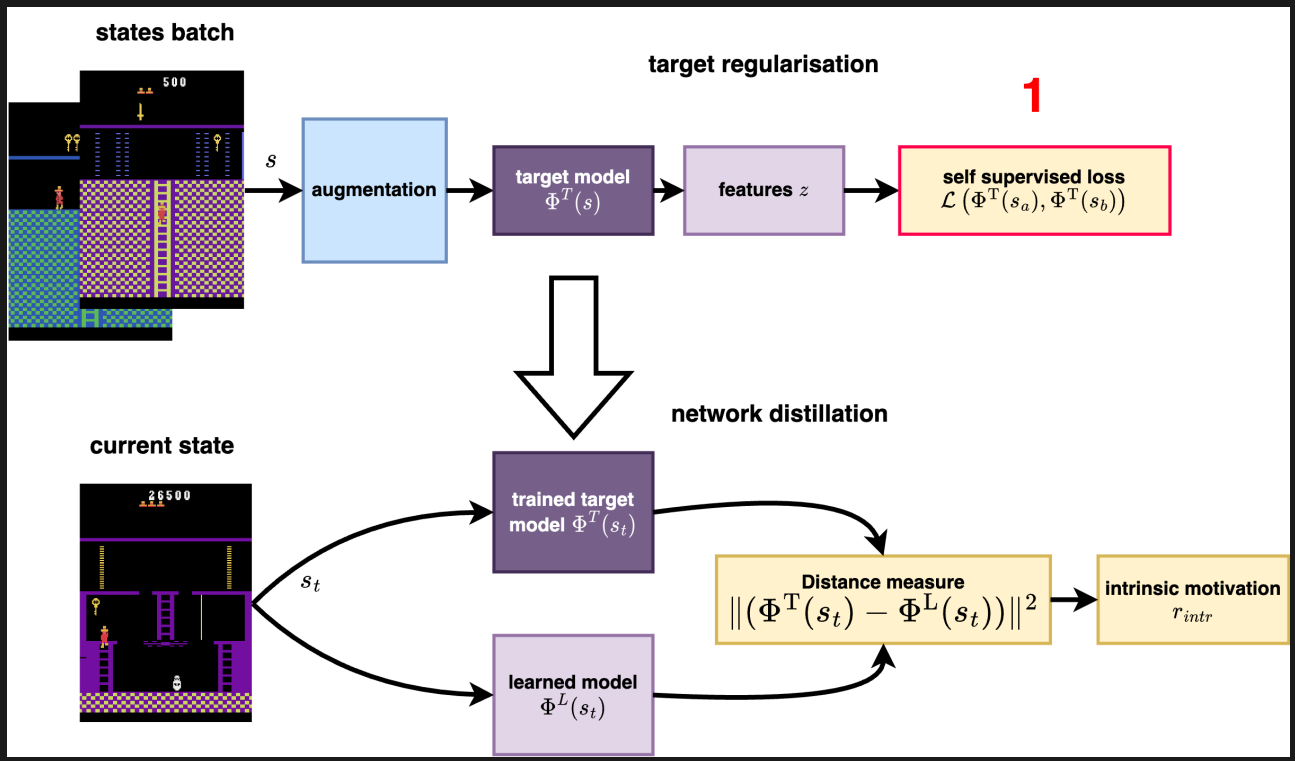
Exploration by self supervised-exploitation

- we extended existing idea of Random Network Distillation



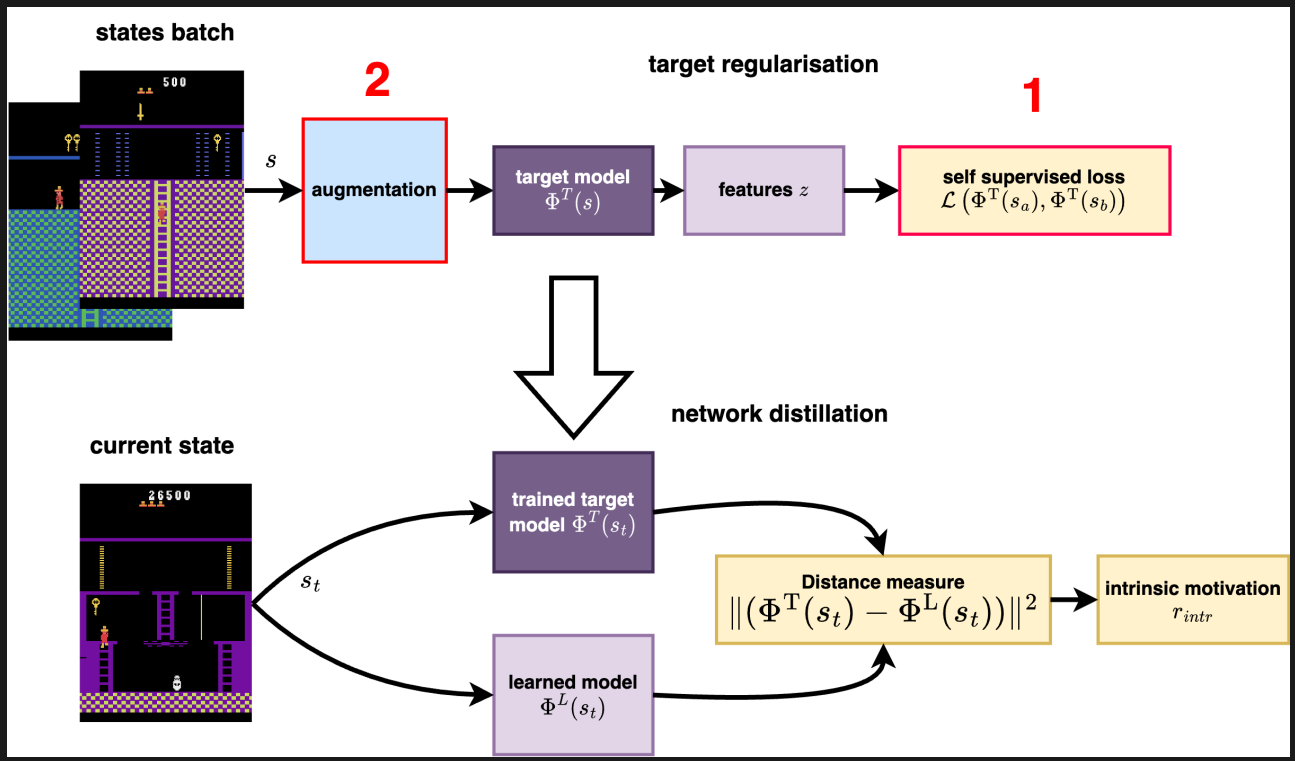
Exploration by self supervised-exploitation

- we extended existing idea of Random Network Distillation
- 1 : for target model, self supervised training is used



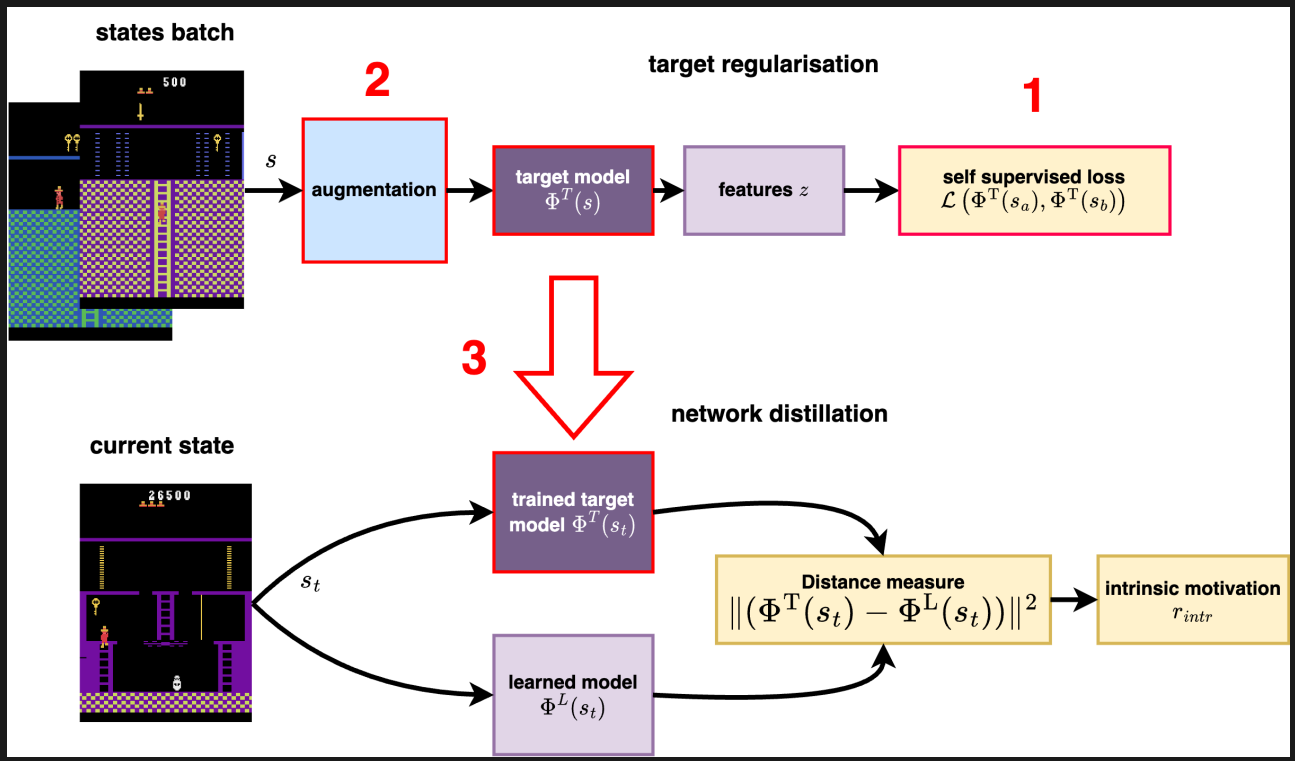
Exploration by self supervised-exploitation

- we extended existing idea of Random Network Distillation
- 1 : for target model, self supervised training is used
- 2 : augmented states are used to train target model



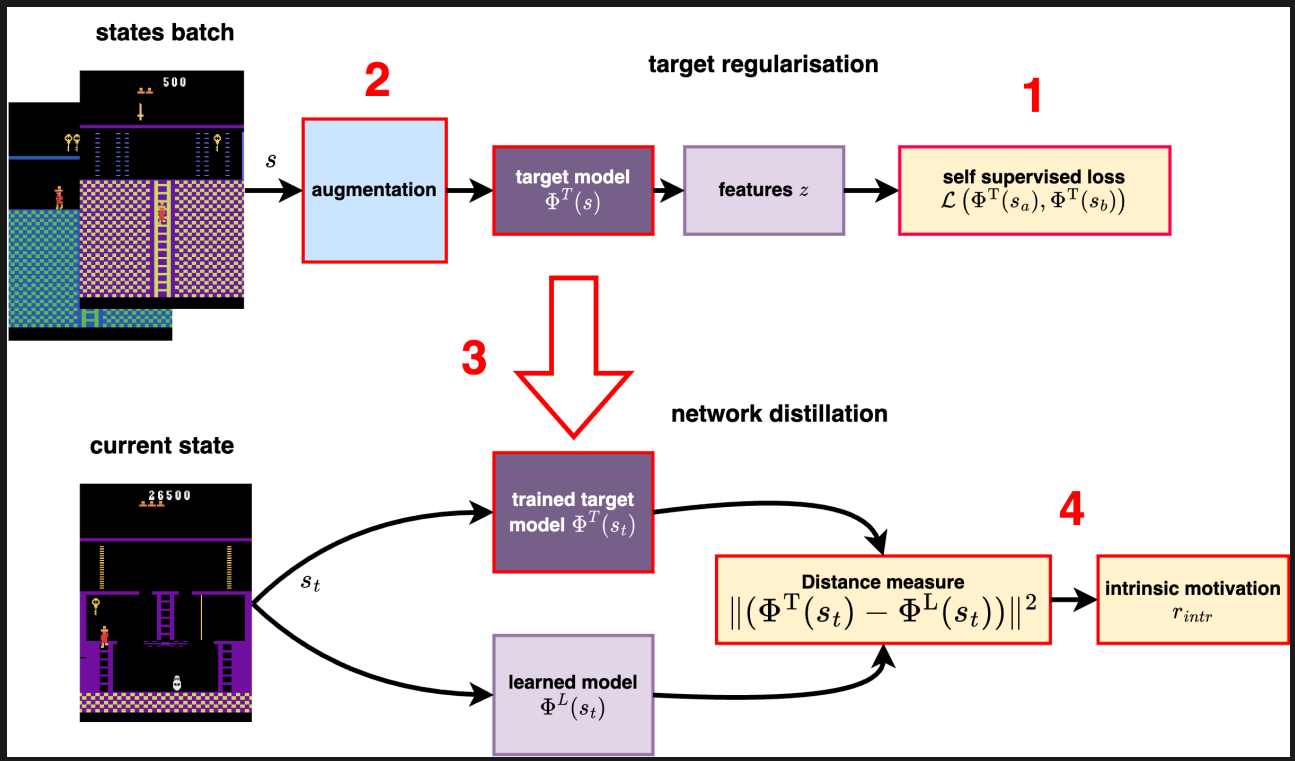
Exploration by self supervised-exploitation

- we extended existing idea of Random Network Distillation
- 1 : for target model, self supervised training is used
- 2 : augmented states are used to train target model
- 3 : target model is used as distillation source



Exploration by self supervised-exploitation

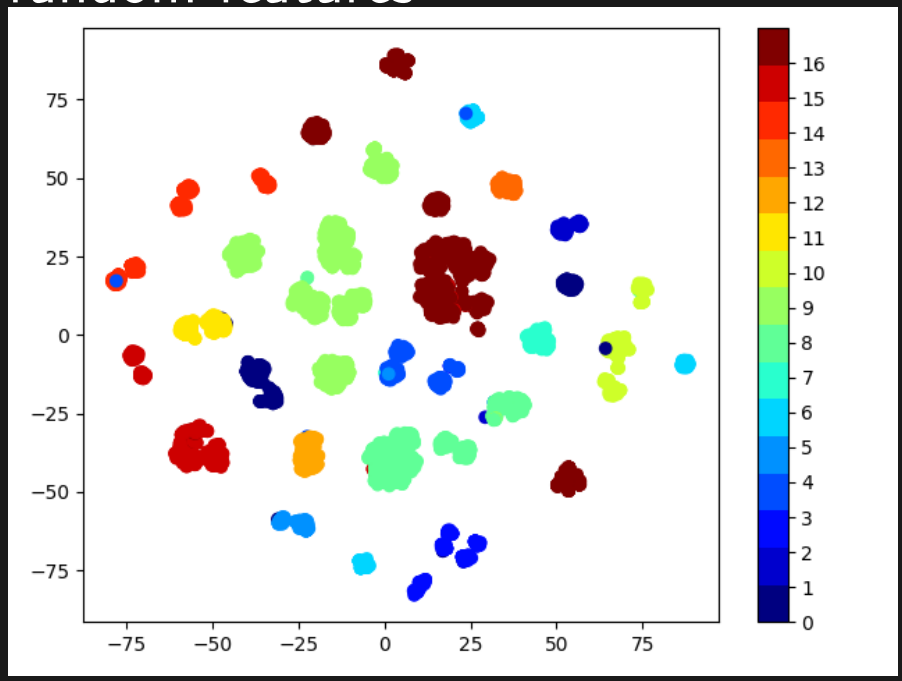
- we extended existing idea of Random Network Distillation
- 1 : for target model, self supervised training is used
- 2 : augmented states are used to train target model
- 3 : target model is used as distillation source
- 4 : distillation error is used for intrinsic motivation



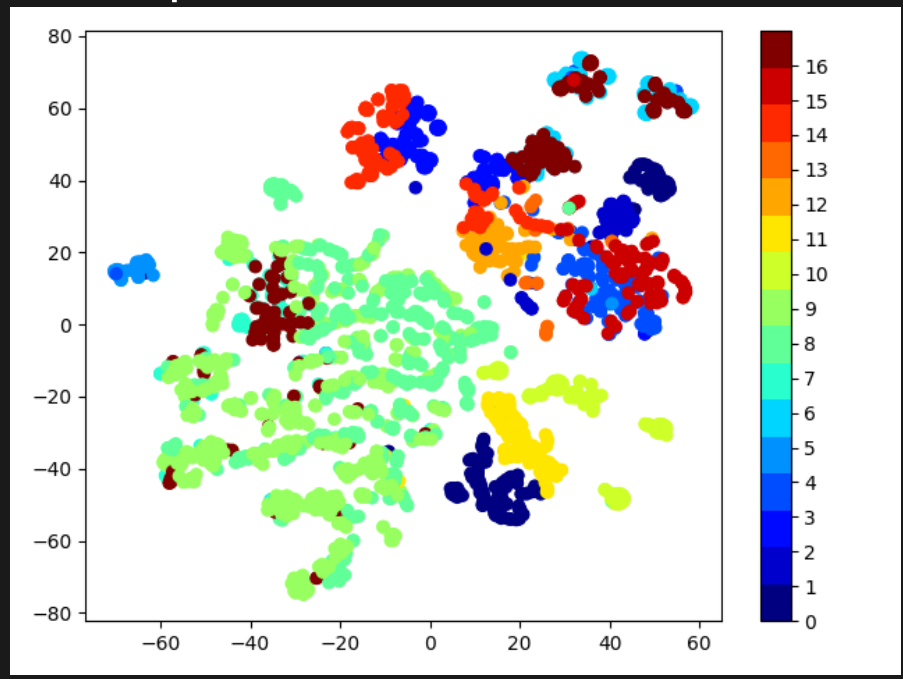
Trained features

- t-SNE features projection for random and trained models

random features



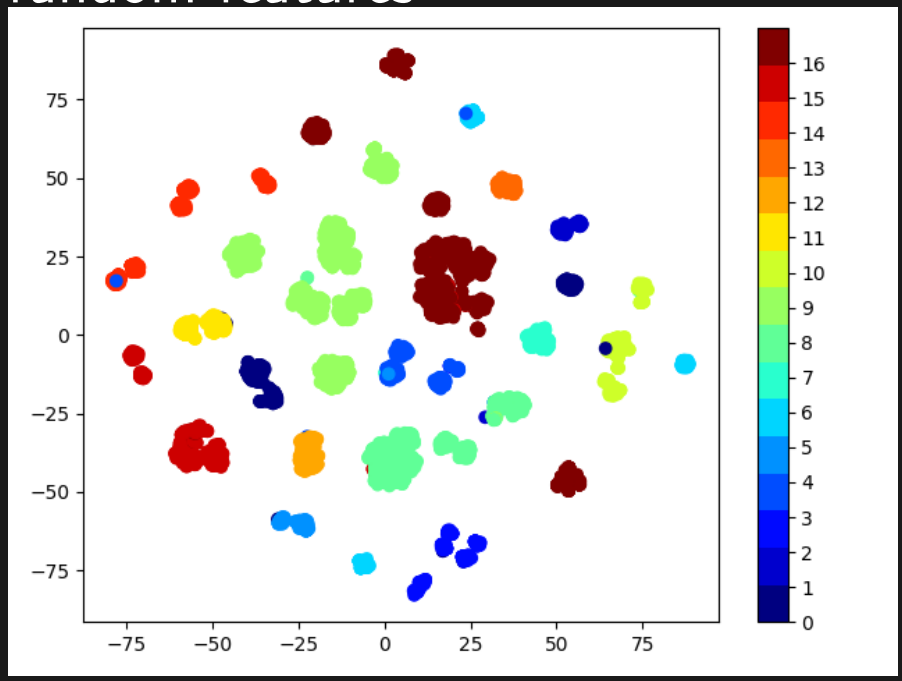
self supervised trained features



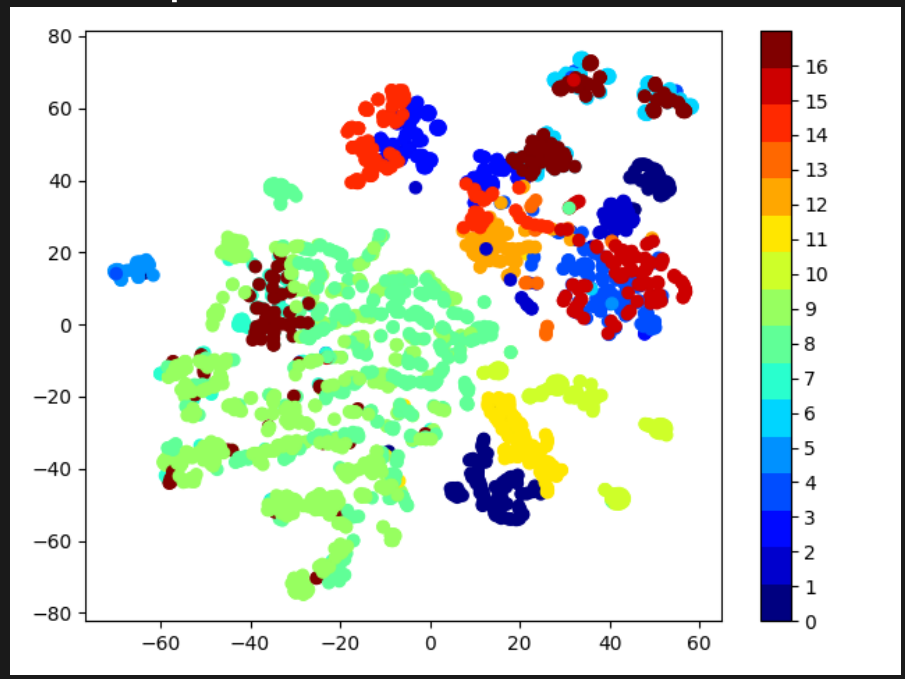
Trained features

- t-SNE features projection for random and trained models
- color represents different rooms in Atari Montezuma's Revenge

random features



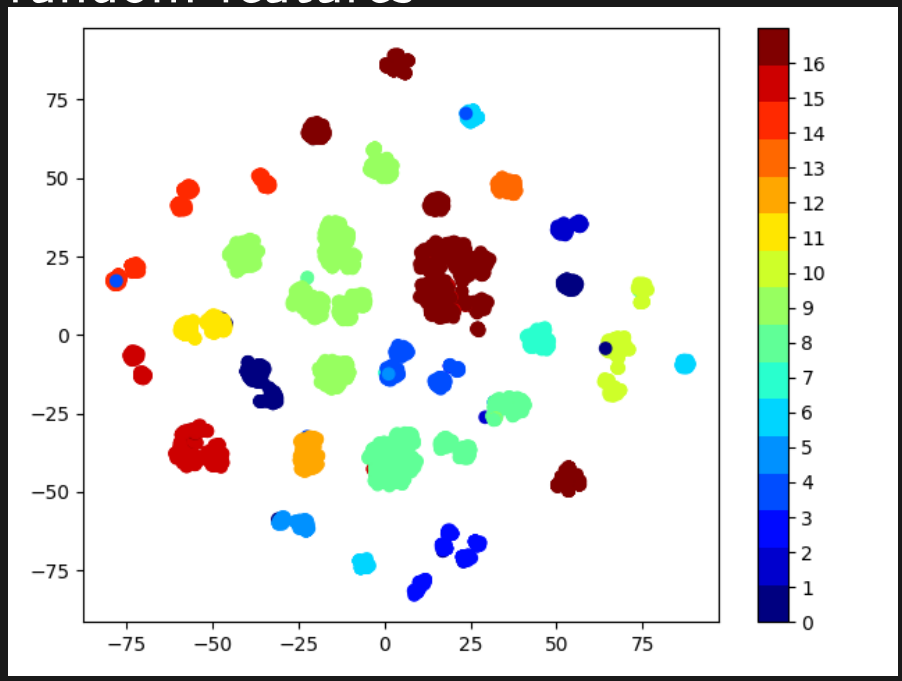
self supervised trained features



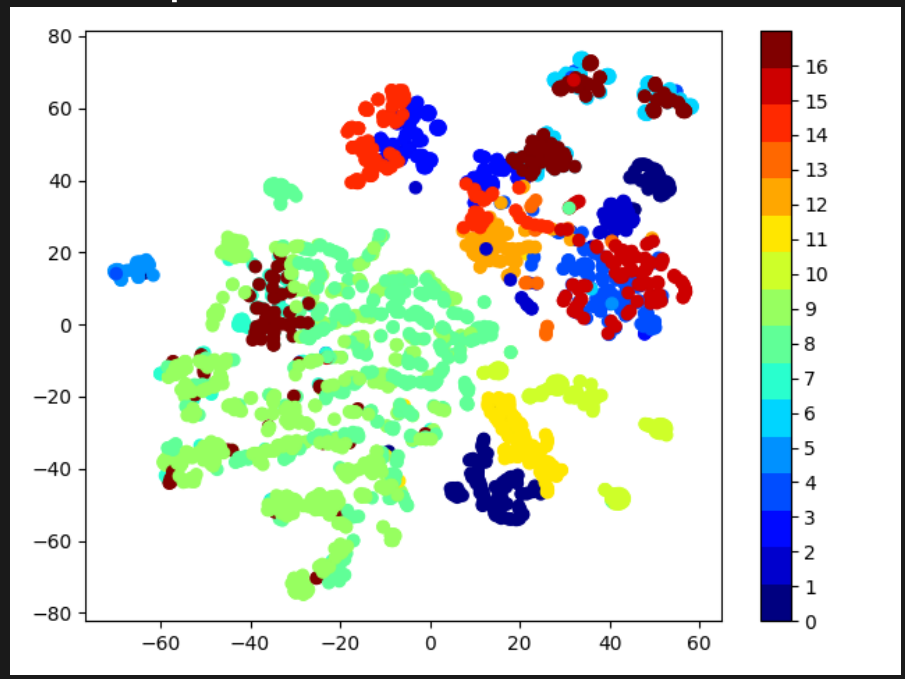
Trained features

- t-SNE features projection for random and trained models
- color represents different rooms in Atari Montezuma's Revenge
- self supervised features provides much bigger variance
- preventing agent to stuck

random features



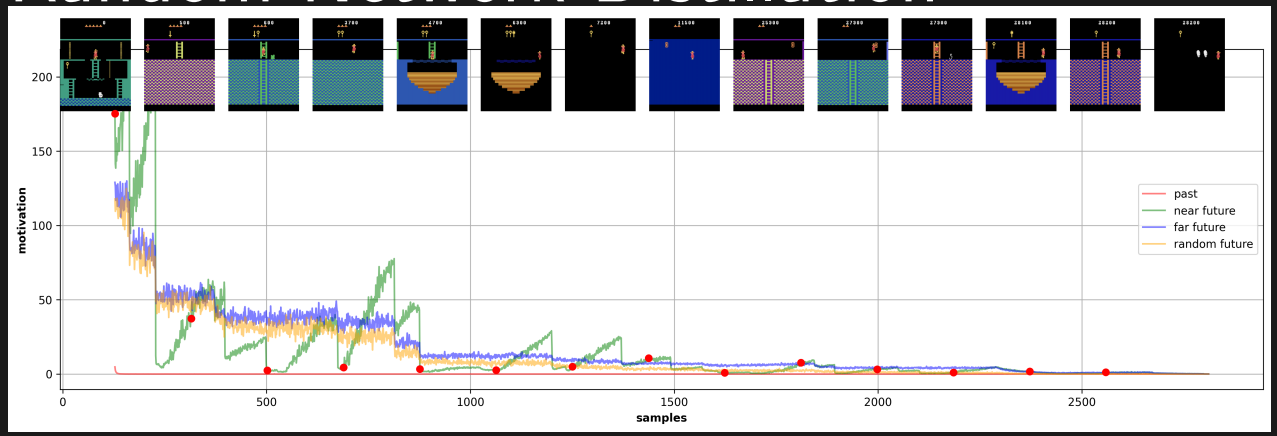
self supervised trained features



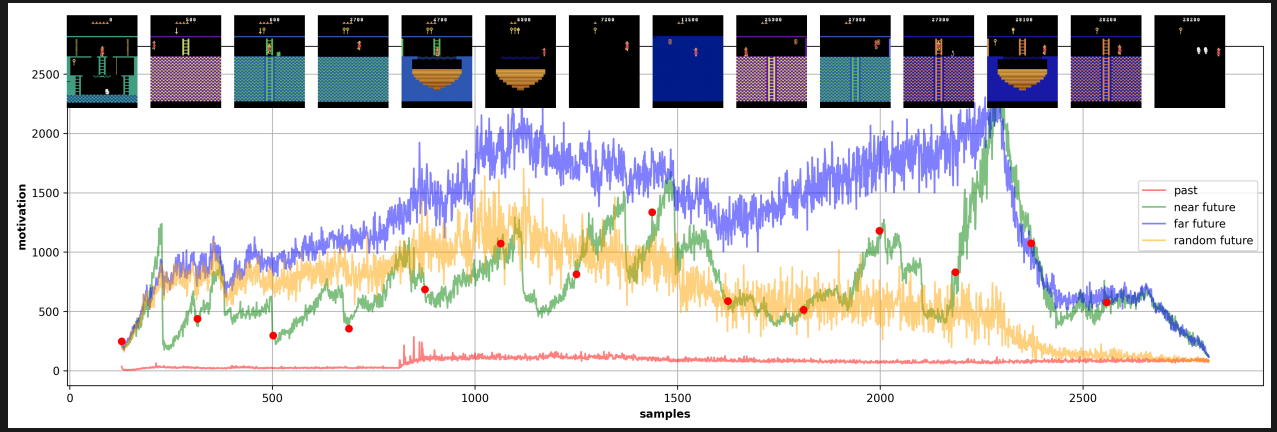
Exploration signal

- Random Network Distillation signal decrease over time

Random Network Distillation



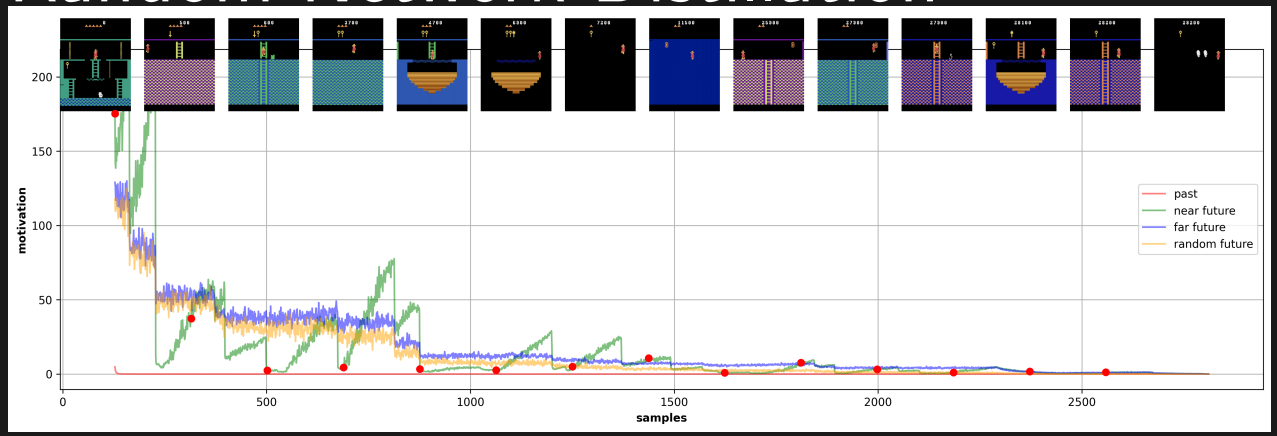
our method



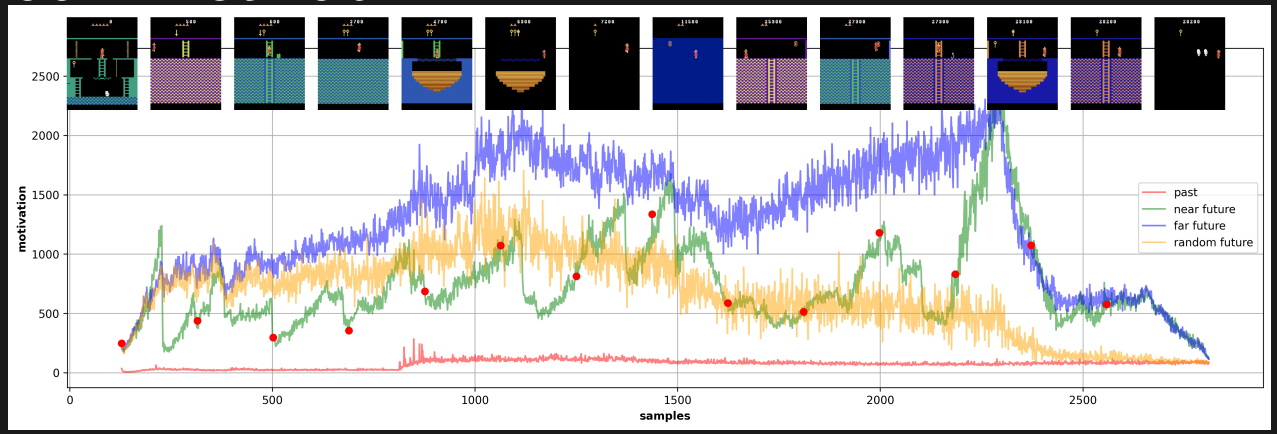
Exploration signal

- Random Network Distillation signal decrease over time
- our method provides more informative signal

Random Network Distillation

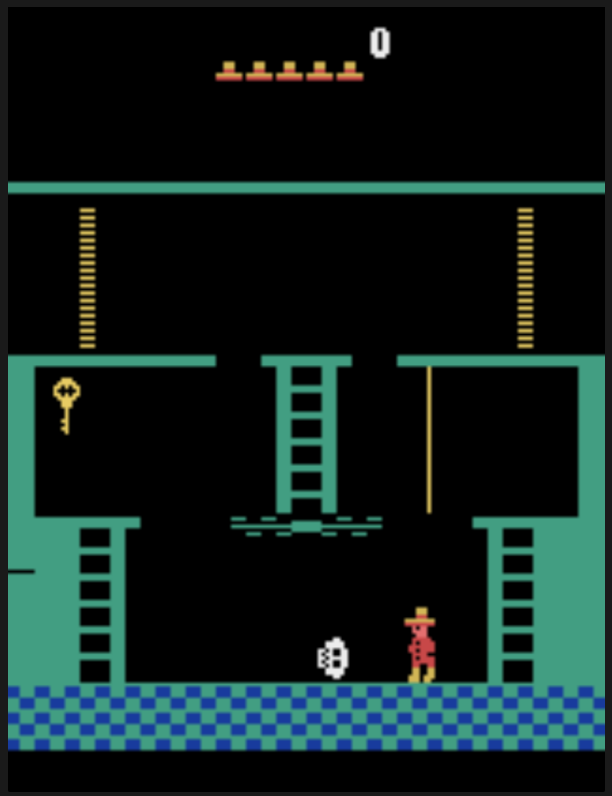


our method



Results

- Montezuma's Revenge, with score 25 000+
- Private Eye, with score 12 000+
- Venture, Gravitar
- 128M samples total - only single GPU needed

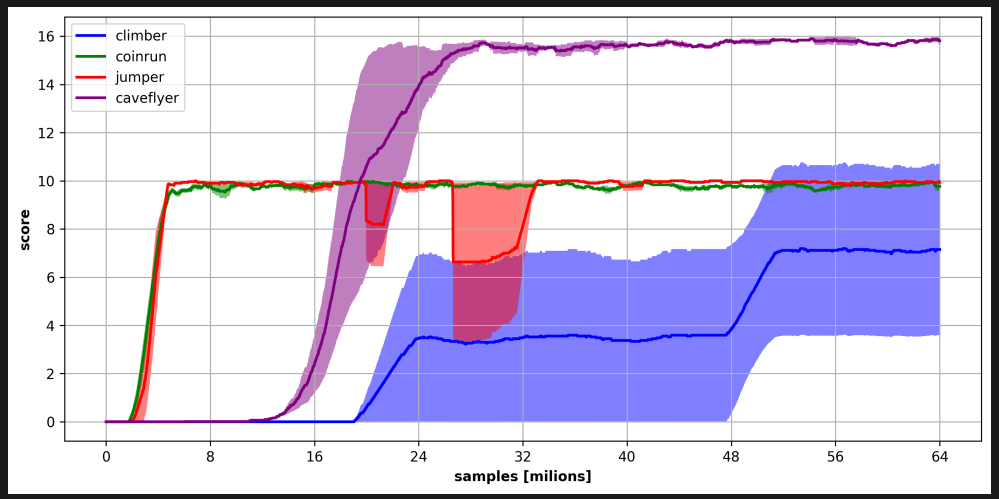


video of playing agent

Results

solved Procgen hard exploration seeds

environments : Caveflyer, Climber, Coinrun, Jumper



video of playing agent

Summary

- New sample efficient exploration method
- Novel view to features z-spaces
- solved environments :
  Procgen (Climber, Caveflyer, Coinrun, Jumper)
  Atari (Montezuma, Solaris, Private Eye, Venture, Gravitar)
- github sources :
  `https://github.com/Iskandor/MotivationModels`
  `https://github.com/michalnand/reinforcement_learning`