LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Hidden Attacks in Multi-Agent Reinforcement Learning

Aufgabensteller:

Prof. Dr. Claudia Linnhoff-Popien

Betreuer:

Thomy Phan

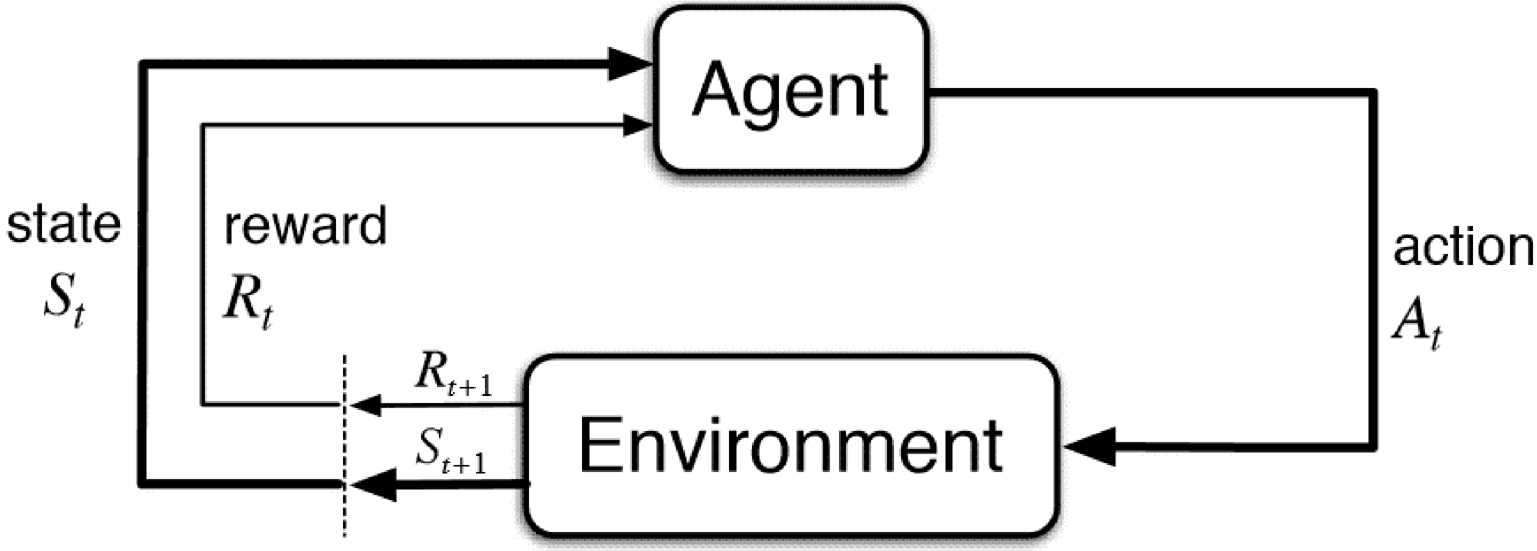
Philipp Altmann

Student:

Arnold Unterauer

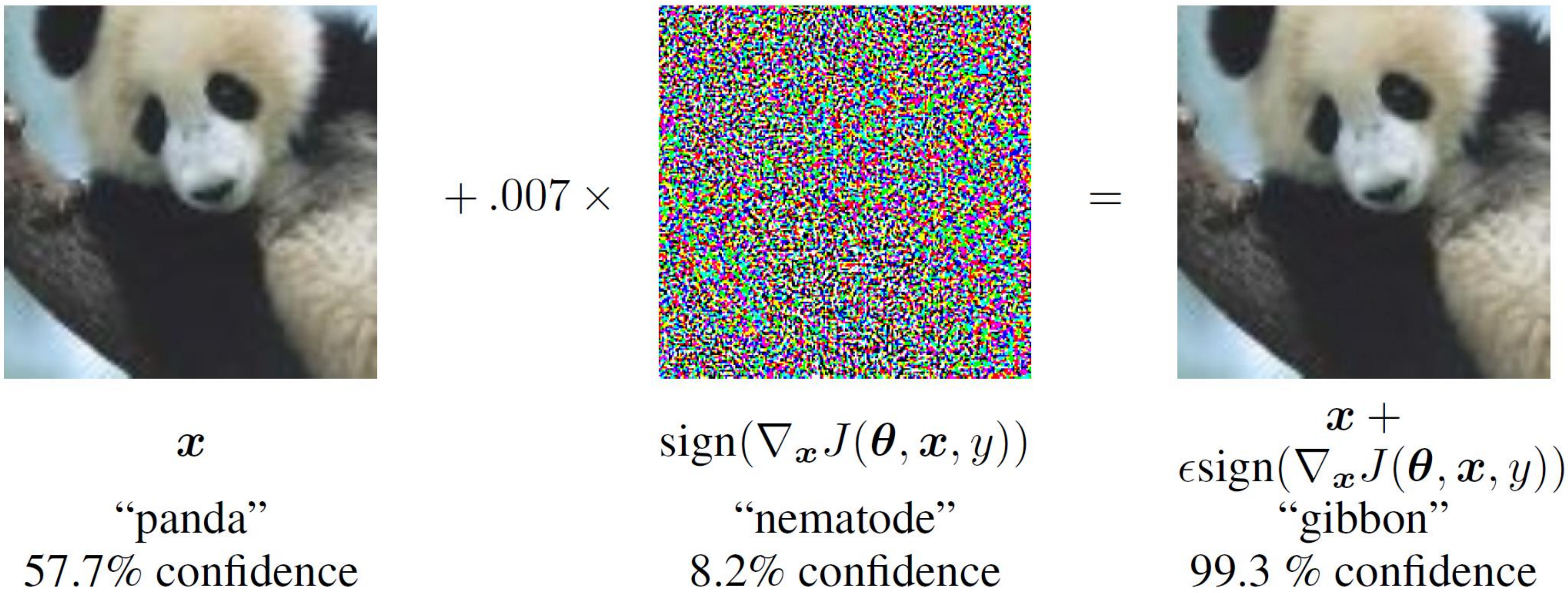
Motivation

Source: https://t.ly/r7wi https://t.ly/msTd https://t.ly/jnk5 https://t.ly/iAX6E



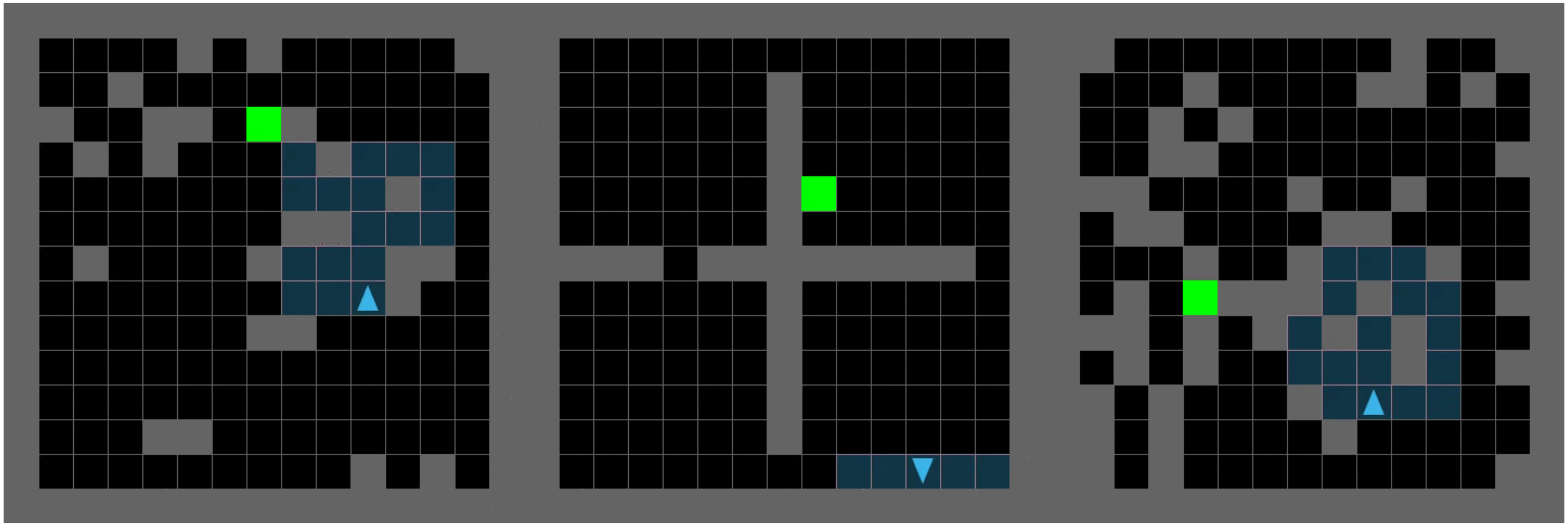
Source: Richard S. Sutton & Andrew G. Barto (1998) **Reinforcement Learning: An Introduction**

Adversarial Attacks



Adversarial Attack **Natural Example**

Source: Goodfellow et al. (2014) **Explaining and Harnessing Adversarial Examples**



Training on randomized environments

Adversarial Attack

Source: Dennis et al. (2021) **Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design**

Agents can fail at any time

⇓

Agents can be malicious

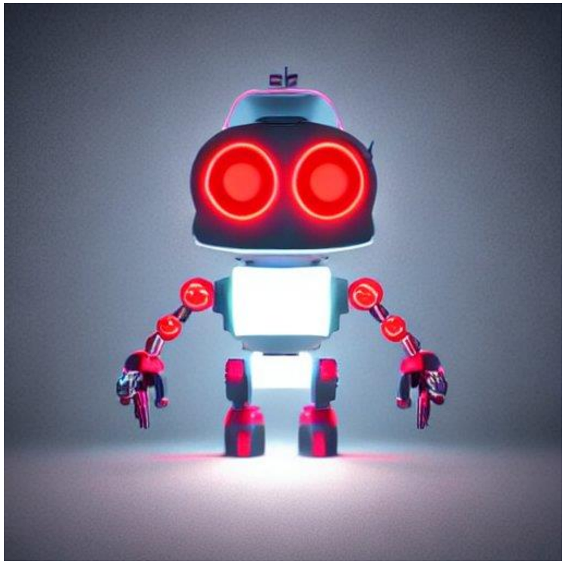
⇓

Agents can be **secretly** malicious

How to work with such agents?

How to create such an agent?

malicious intent

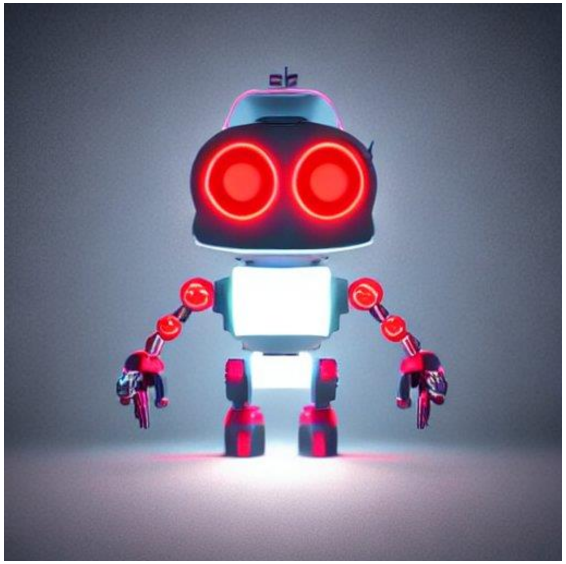


undetected

ISAAC Invader

Source: **Stable Diffusion** Prompt: „Evil red Bot with angry eyes and light, Illustration“

Antagonistic Behavior *



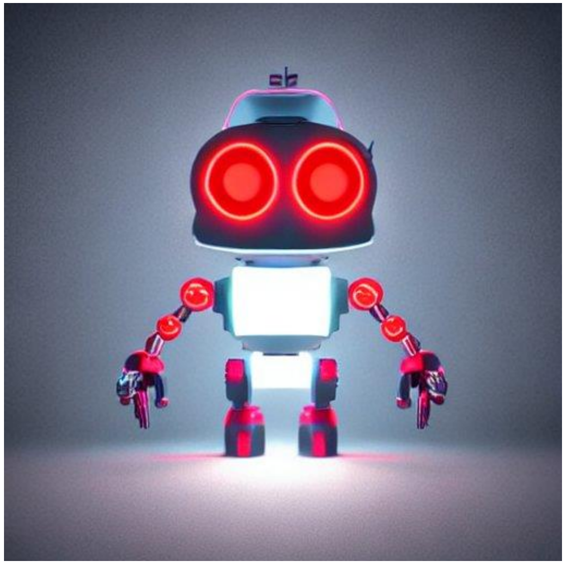
Protagonist Imitation

ISAAC Invader

* Phan et al. (2020) **Learning and Testing Resilience in Cooperative Multi-Agent Systems**

Infiltrating Stealth Agent Attack Controller (ISAAC)

Antagonistic Behavior *



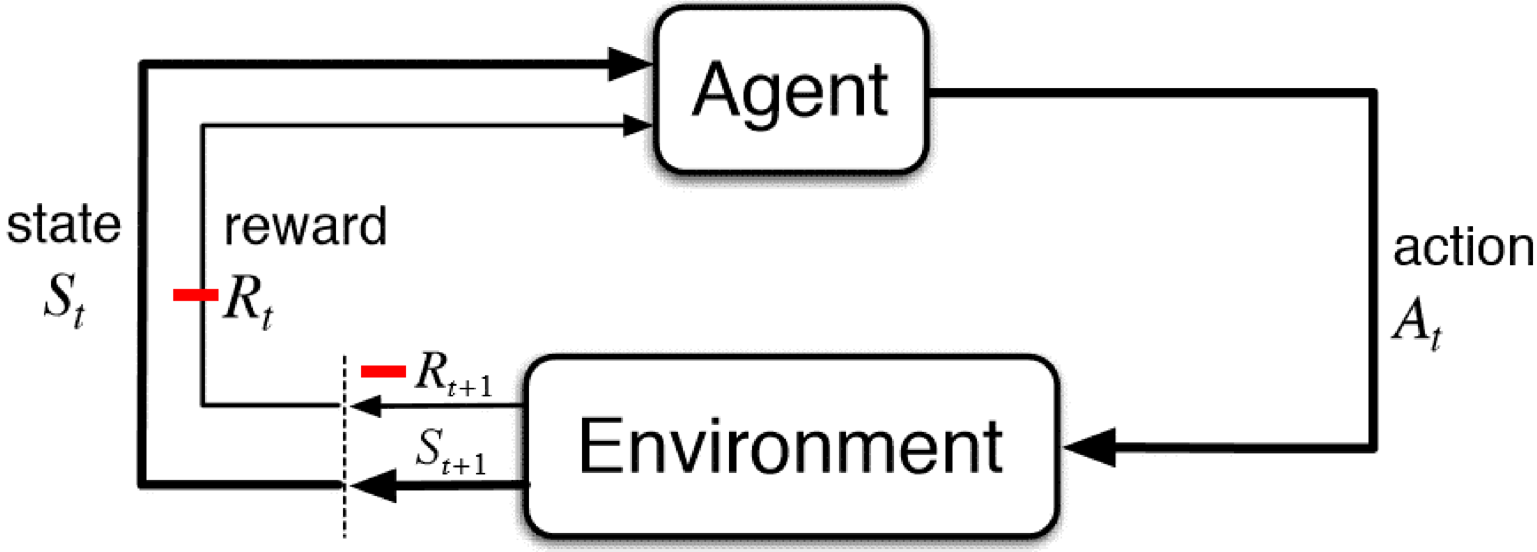
ISAAC Invader

Imitate Behavior

Mimic State Features

* Phan et al. (2020) **Learning and Testing Resilience in Cooperative Multi-Agent Systems**

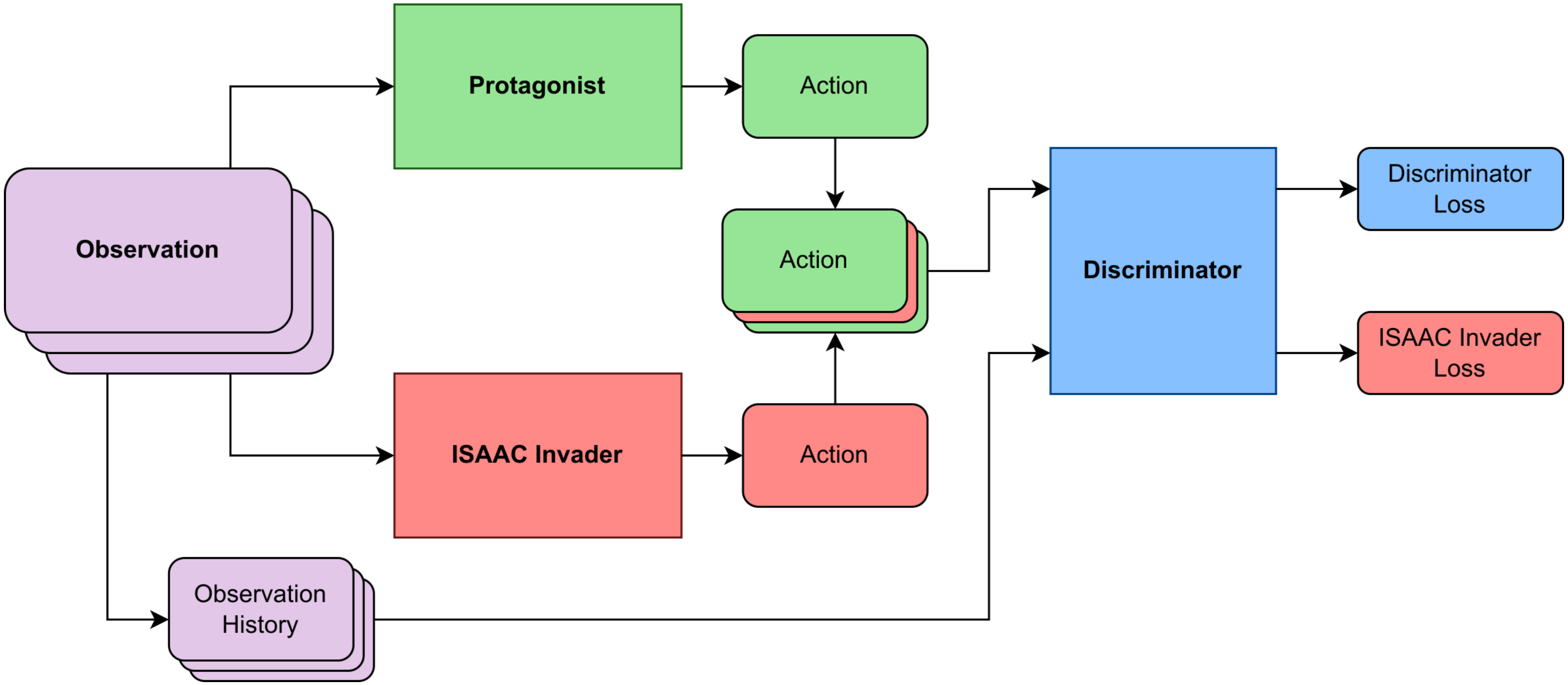
Antagonistic Behavior (Naive Invader)



$$R_{ISAAC} = -R_{pro}$$

Source: Phan et al. (2020) **Learning and Testing Resilience in Cooperative Multi-Agent Systems**

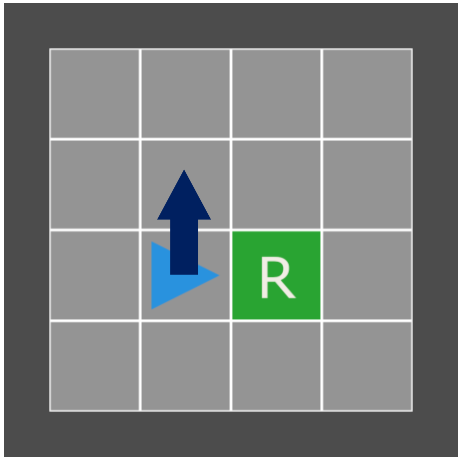
Imitate Behavior

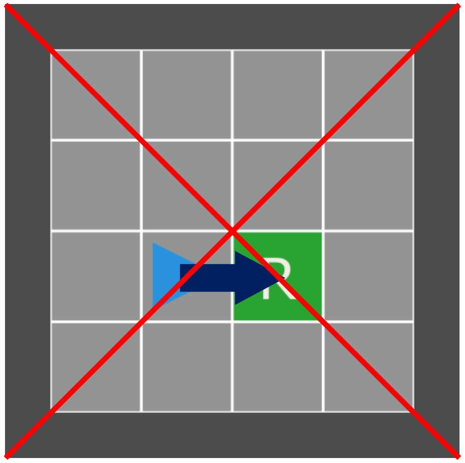


$$H_p(q) = -\frac{1}{T}\sum_{t=1}^{T}[y_{v_t} * \log(p(y_{v_t})) + (1 - y_{v_t}) * \log\left(1 - p(y_{v_t})\right)], \quad v_t = (u_t, \tau_t)$$

Source: Rubinstein (1999) **The Cross-Entropy Method for Combinatorial and Continuous Optimization**

Behavior Shaping





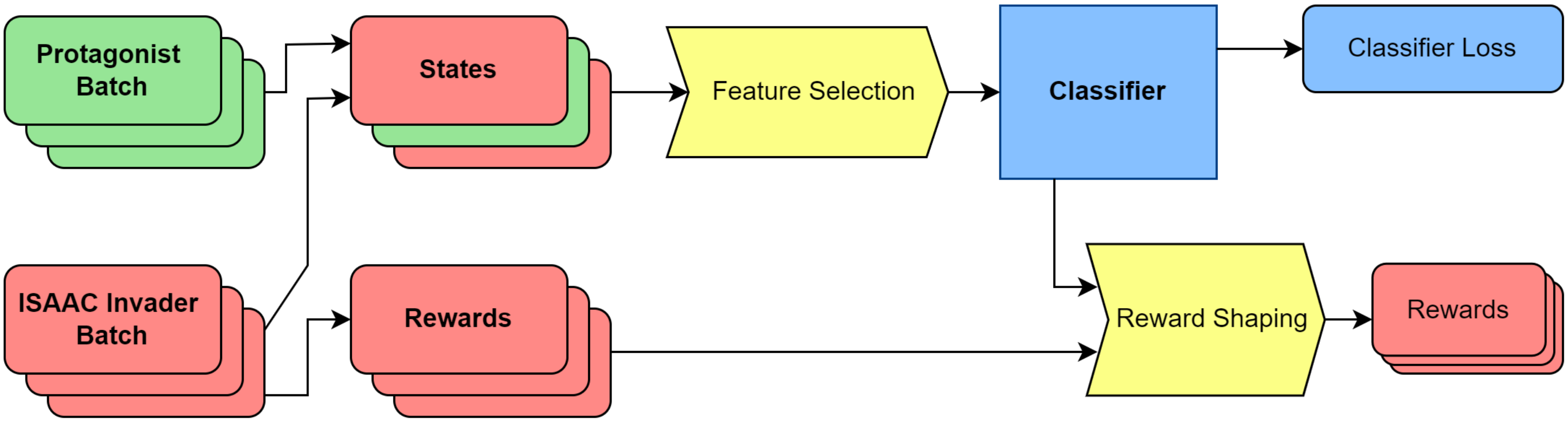
$$H_{bs,p}(q,R) = \begin{cases} H_p(q), & if\ R \leq T_{bs} \\ 0, & otherwise \end{cases}$$

Mimic State Features

Behavior imitation is not enough

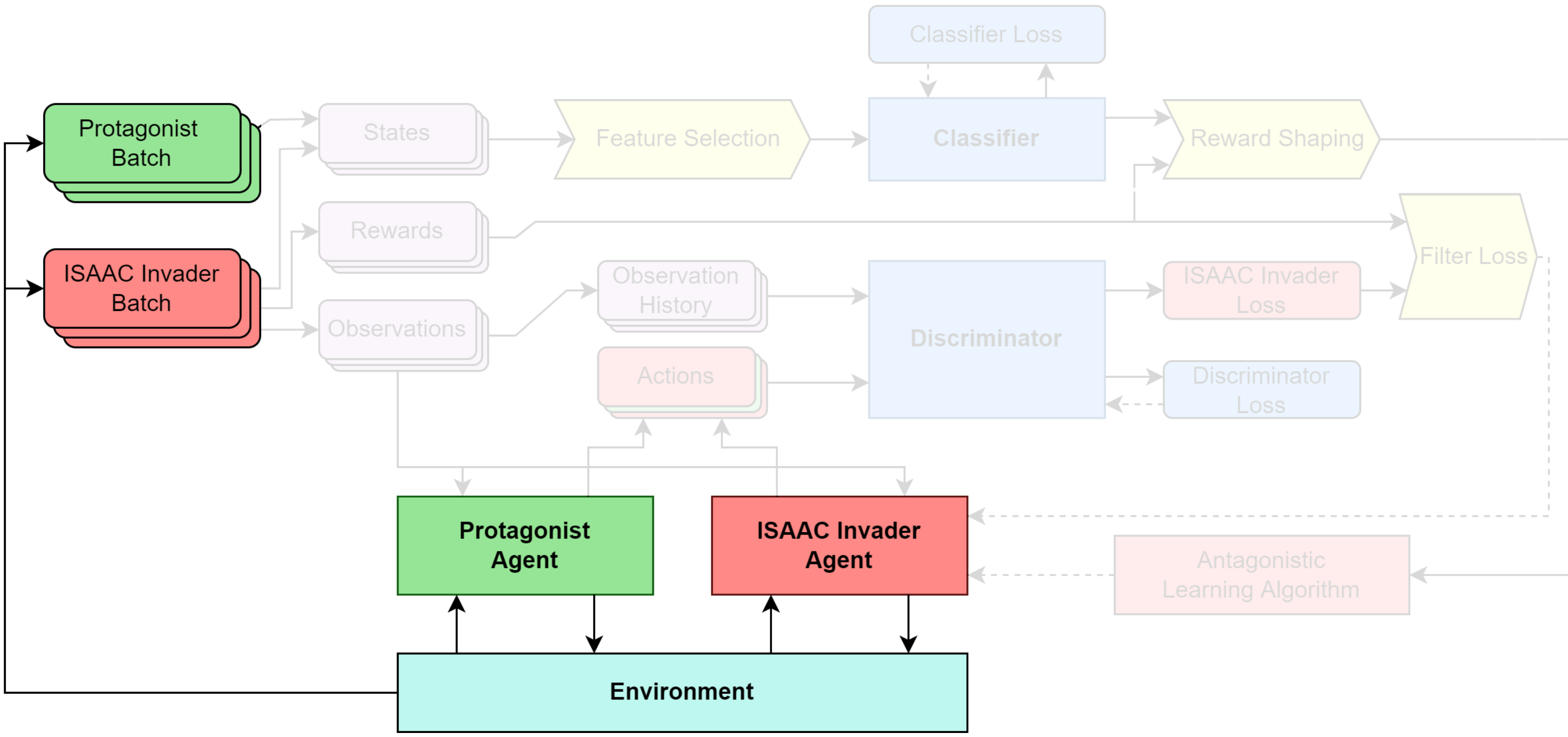
Source: https://johngushue.typepad.com/blog/2011/09/stormtrooper-office-pranks.html

Mimic State Features



$$r\left(p(y_{s'})\right) = \begin{cases} r + F & \text{for} \quad p(y_{s'}) \geq 0.5 \\ r - F & \text{for} \quad p(y_{s'}) < 0.5 \end{cases}$$

Imitate Behavior + Mimic State Features + Antagonistic Behavior



3 Stalkers vs 3 Zealots (3s_vs_3z)

Homogeneous & asymmetric

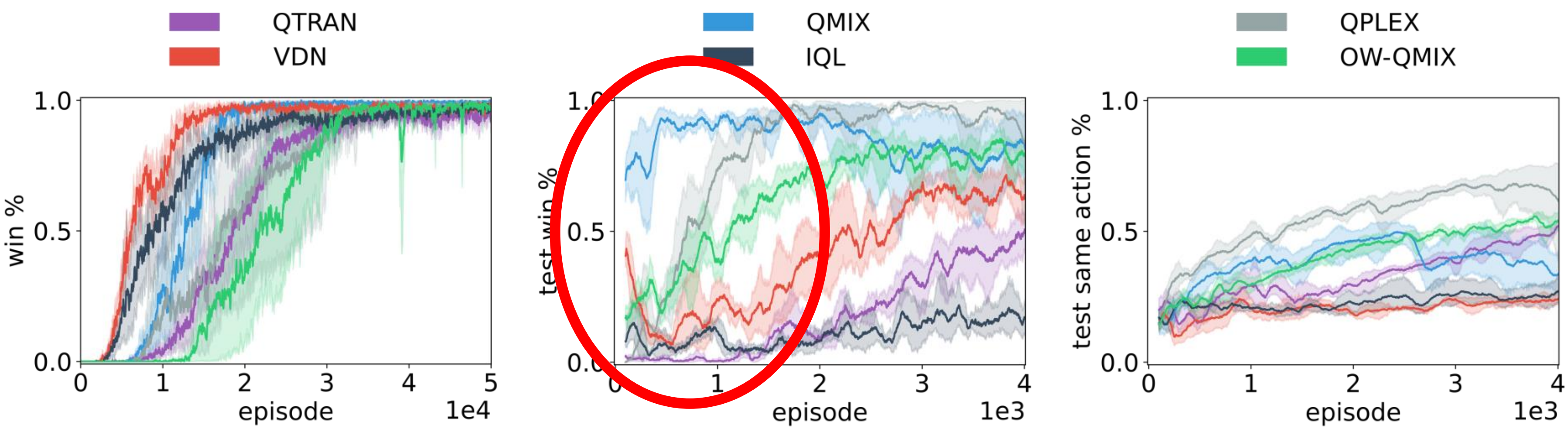
Micro-Trick: Kiting

Difficulty: Easy

Interesting Aspect: Positioning

Source: Samvelyan et al. (2019) **The StarCraft Multi-Agent Challenge**

3 Stalkers vs 3 Zealots - Training



Protagonist Training

ISAAC Invader Training

Random Invader can already lower the performance

With learning progress the win rate and action matching also increases

After training the win rates are still lower than the protagonist

3 Stalkers vs 3 Zealots - Evaluation

3s_vs_3z	No Substitution	ISAAC Invader	Naive Invader	Random Invader
QTRAN	0.82 ± 0.01	0.56 ± 0.01	0.01 ± 0.00	0.13 ± 0.01
QMIX	1.00 ± 0.00	0.82 ± 0.01	0.68 ± 0.01	0.90 ± 0.01
QPLEX	1.00 ± 0.00	0.84 ± 0.01	0.00 ± 0.00	0.04 ± 0.01
VDN	0.99 ± 0.00	0.75 ± 0.01	0.20 ± 0.01	0.53 ± 0.01
IQL	0.96 ± 0.01	0.27 ± 0.01	0.07 ± 0.01	0.23 ± 0.01
OW-QMIX	0.98 ± 0.00	0.79 ± 0.01	0.21 ± 0.01	0.31 ± 0.01

Win rates

3s_vs_3z	No Substitution	ISAAC Invader	Naive Invader	Random Invader
QTRAN	1.00 ± 0.00	0.56 ± 0.00	0.23 ± 0.00	0.16 ± 0.00
QMIX	1.00 ± 0.00	0.34 ± 0.00	0.08 ± 0.00	0.15 ± 0.00
QPLEX	1.00 ± 0.00	0.66 ± 0.00	0.16 ± 0.01	0.15 ± 0.00
VDN	1.00 ± 0.00	0.26 ± 0.00	0.20 ± 0.00	0.15 ± 0.00
IQL	1.00 ± 0.00	0.24 ± 0.00	0.12 ± 0.00	0.16 ± 0.00
OW-QMIX	1.00 ± 0.00	0.53 ± 0.00	0.10 ± 0.00	0.16 ± 0.00

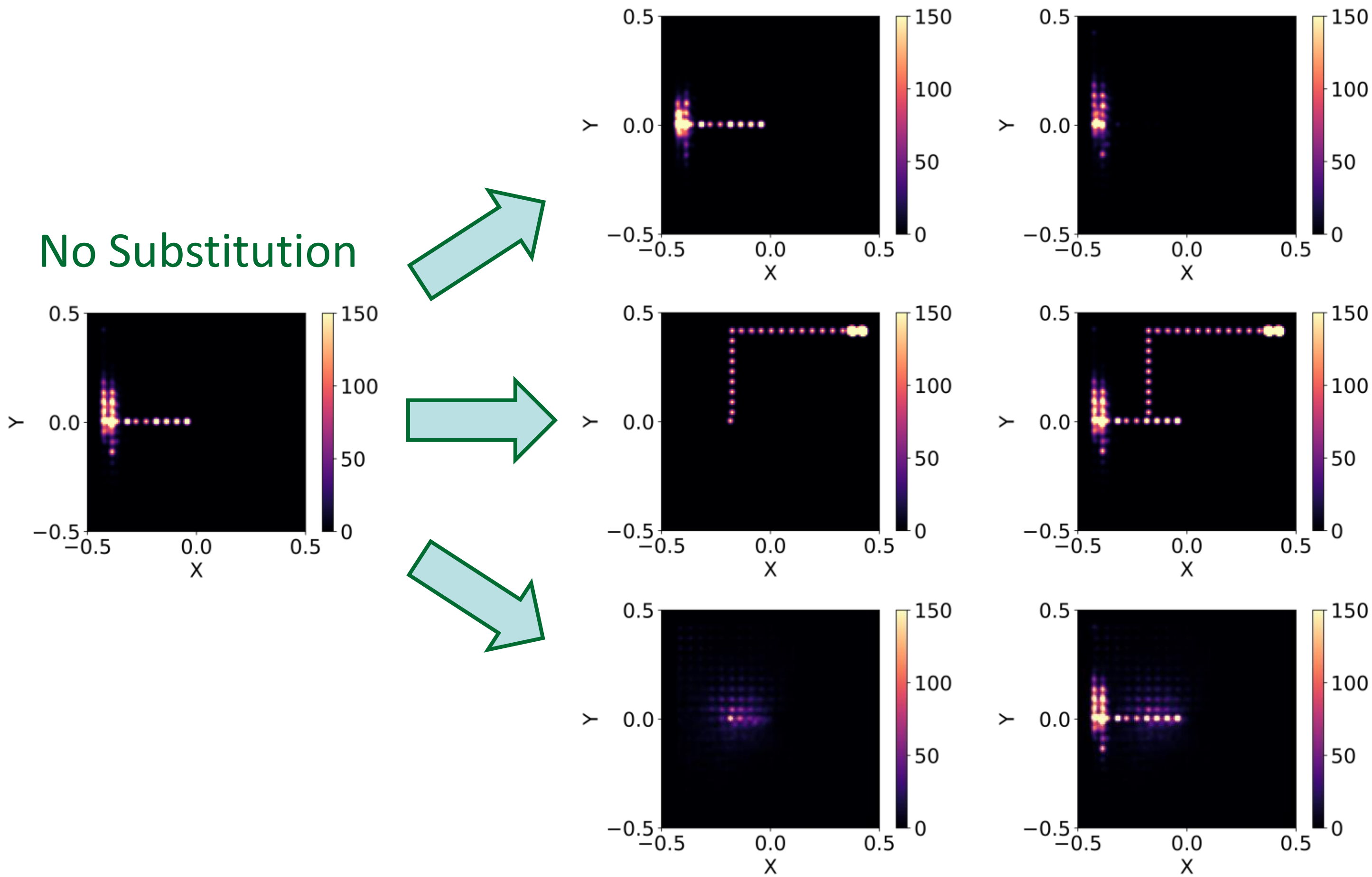
Action Matching

3 Stalkers vs 3 Zealots - Positioning

Invader

Difference

No Substitution



ISAAC Invader
$49.44 \pm 20.99\%$

Naive Invader
$1.31 \pm 2.16\%$

Random Invader
$4.76 \pm 0.20\%$

5 Marines vs 6 Marines (5m_vs_6m)

Homogeneous & asymmetric

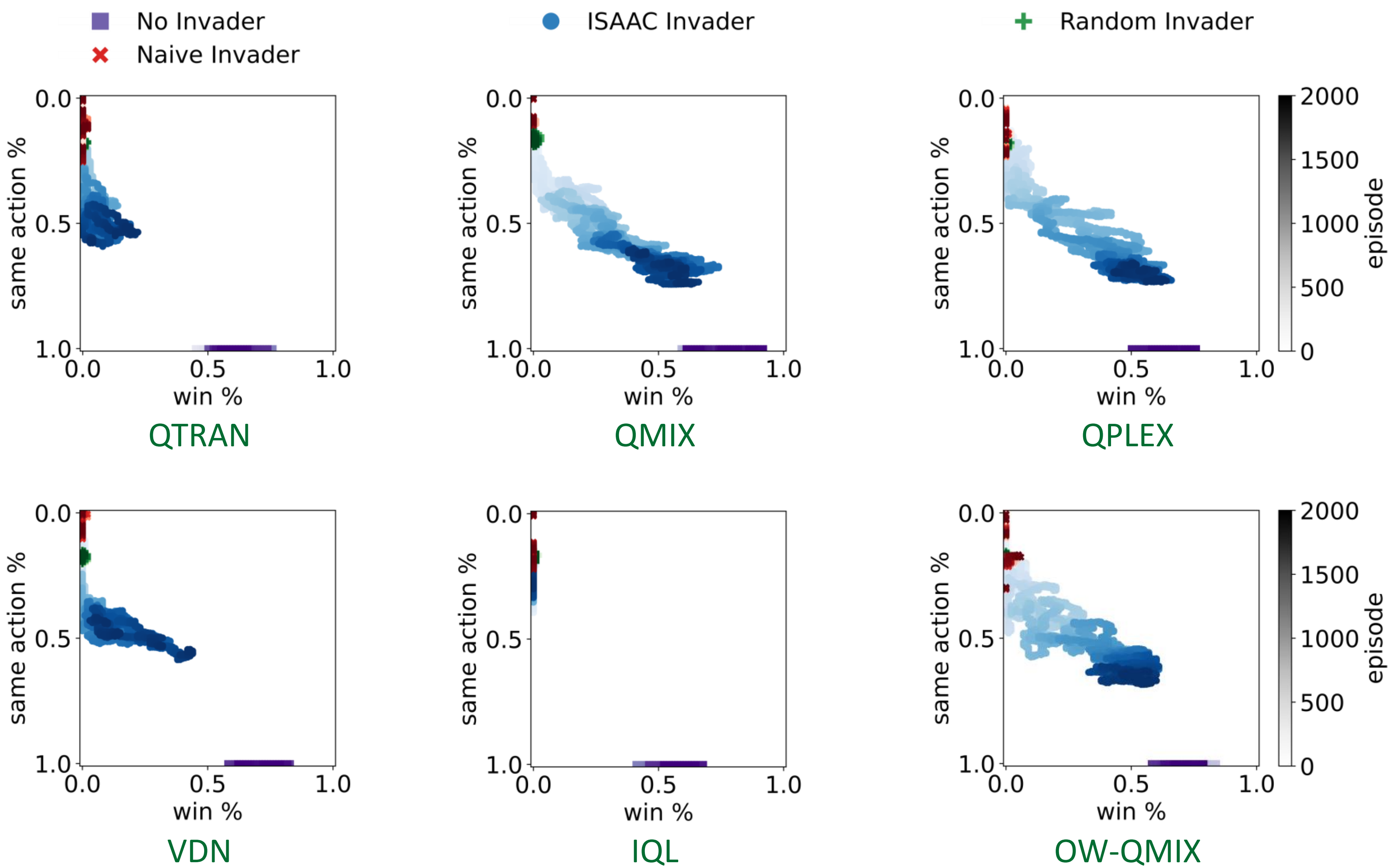
Micro-Trick: Focus Fire

Difficulty: Hard

Interesting Aspect: Difficulty

Source: Samvelyan et al. (2019) **The StarCraft Multi-Agent Challenge**

5 Marines vs 6 Marines - Trade-off



5 Marines vs 6 Marines - Survey

Survey 22 P | 5m_vs_6m	No	Random	Naive	ISAAC	ISAAC	ISAAC	ISAAC	ISAAC
Same action rate (mean)	1.00	0.16	0.04	0.29	0.45	0.54	0.59	0.69
Winrate (mean)	0.76	0.00	0.00	0.02	0.14	0.28	0.35	0.56
Invader id. present	10	22	16	21	6	9	11	16
Correct invader id.	0	21	13	21	2	2	3	4
Protagonists as invader id.	21	4	7	6	10	16	14	30
Pro. as invader id. per ag.	4.2	1	1.75	1.5	2.5	4	3.5	7.5
True Positive Rate	0.58	0.95	0.89	0.93	0.58	0.56	0.68	0.53
True Negative Rate	–	**0.95**	**0.81**	**1.00**	**0.33**	**0.22**	**0.27**	**0.25**
False Negative Rate	0.42	0.05	0.11	0.07	0.42	0.44	0.32	0.47
False Positive Rate	–	**0.05**	**0.19**	**0.00**	**0.67**	**0.78**	**0.73**	**0.75**

45 % Action Matching sufficient

Win rate decrease by 81.58% while remaining hidden

5 Marines vs 6 Marines - Replay

Random Invader

Naive Invader

ISAAC Invader / Naive Invader

ISAAC Invader

Conclusion

- Introduction of ISAAC, consisting of antagonistic and imitation parts
- ISAAC outperforms Naive and Random Invaders in stealth in terms of positioning and action selection (also reflected in replays)
- ISAAC Invader decreases performance while remaining hidden
- ISAAC is effective against state-of-the-art algorithms and in various environments of the SMAC
- Ability to find a proper trade-off through analysis of training plots
- Limitations in certain environments and algorithms

Thank you for your attention

Future Work

- Substitute more protagonists as Invaders
- Other Reinforcement Learning Algorithms as victims
- Different algorithms for the protagonist and the invader
- Other GAN / classifier variations
- Other dissimilarity measures
- Study social aspects
- Defense mechanism

Hyperparameter & Architecture

- ε-Greedy Method: decay from 1.0 to 0.05 over 100000 time steps
- 2000000 training steps for protagonists
- Batch size 16, total 5000 batches
- RMSprop with $\alpha = 0.99$ and $\varepsilon = 0.00001$
- Learning rate $\alpha = 0.0005$
- Target network parameter θ^{-} update every 200
- Discount factor $\gamma = 0.99$
- RNN Agents with GRU
- Gradient clip at L2 norm of 10

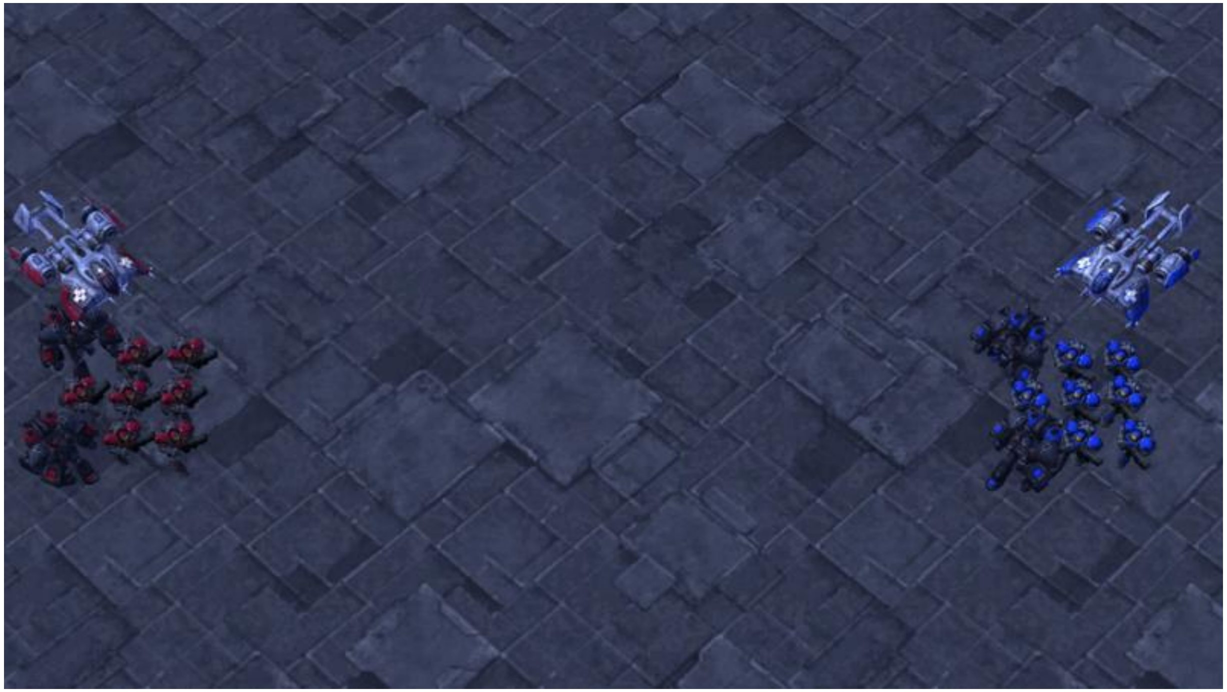
Hyperparameter & Architecture

- Threshold T = 1.0
- Mimic State Classifier Reward F = 0.1
- Discriminator and Classifier use Leaky ReLU with α = 0.25

- Discriminator Action Input: 64 - 32
- Discriminator History Input: 512 - 256 - 256
- Discriminator Combination: 256 - 256 - 64
- Output with Sigmoid function into (0, 1) for the predicted probability

- Classifier: 1024 - 512 - 256 - 64
- Output with Sigmoid function into (0, 1) for the predicted probability

1 Medivac, 2 Marauders & 7 Marines (MMM)



Heterogeneous & symmetric

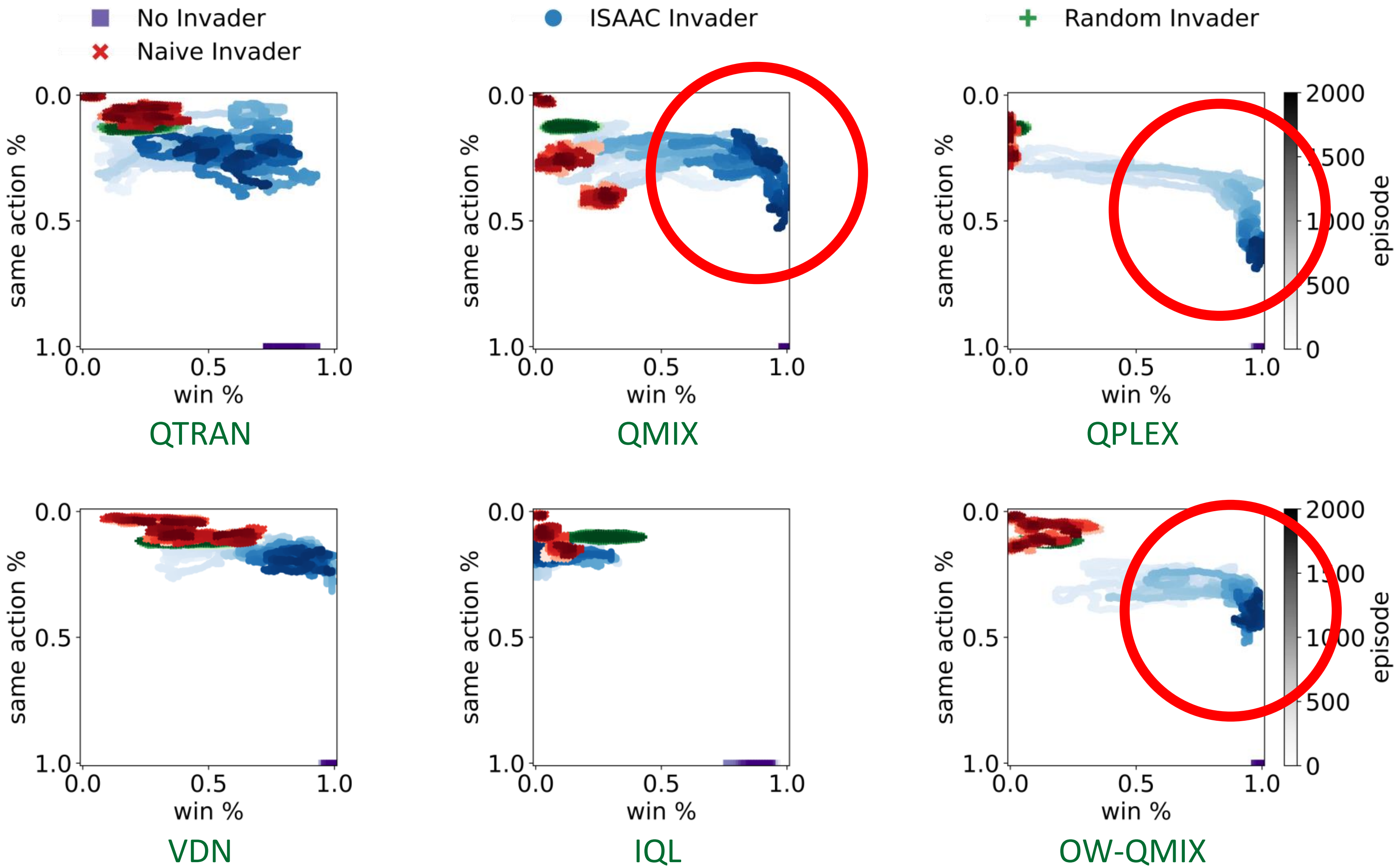
Micro-Trick: Focus Fire

Difficulty: Easy

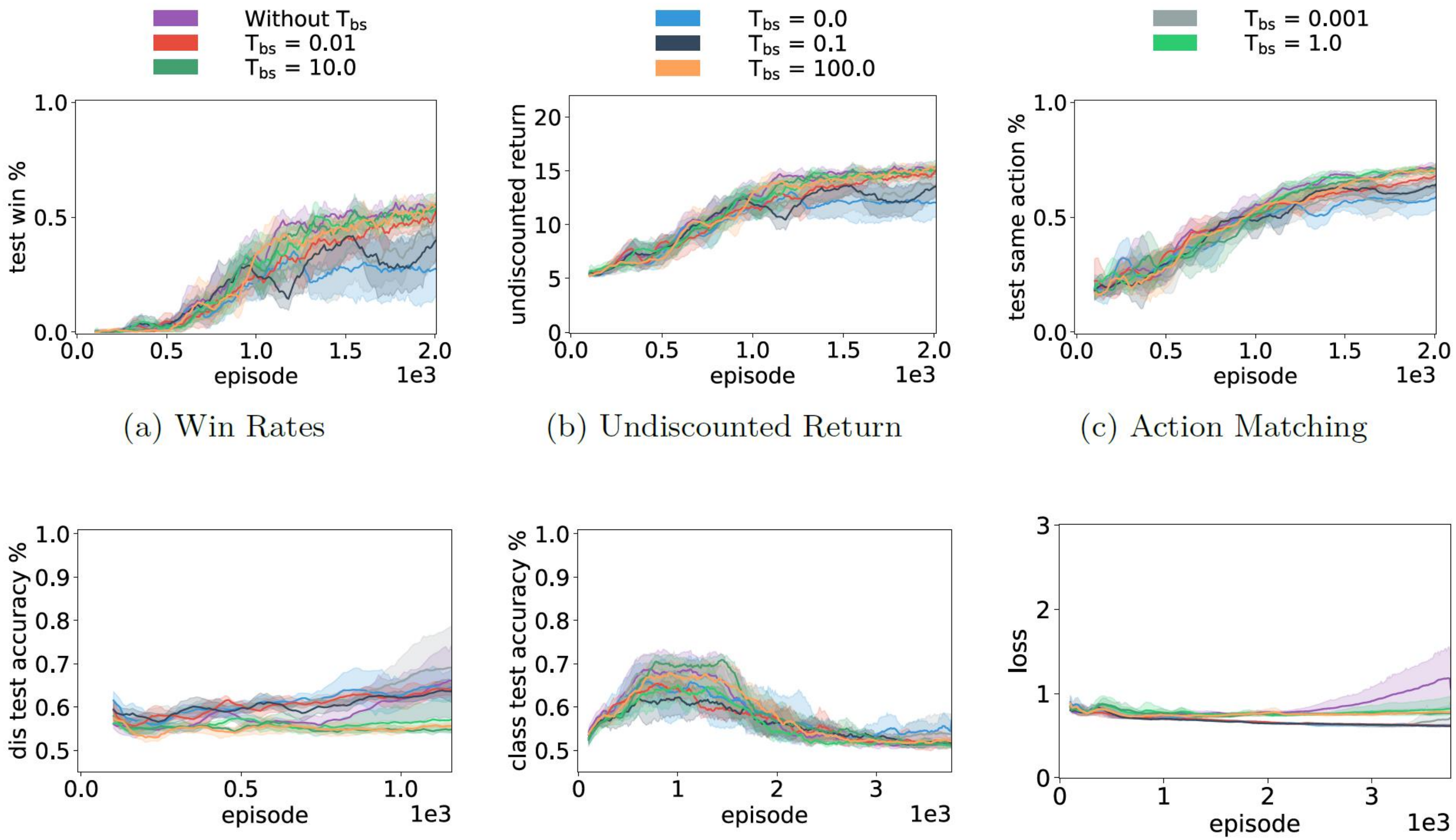
Interesting Aspect: Healing Unit

Source: Samvelyan et al. (2019) **The StarCraft Multi-Agent Challenge**

1 Medivac, 2 Marauders & 7 Marines - Trade-off



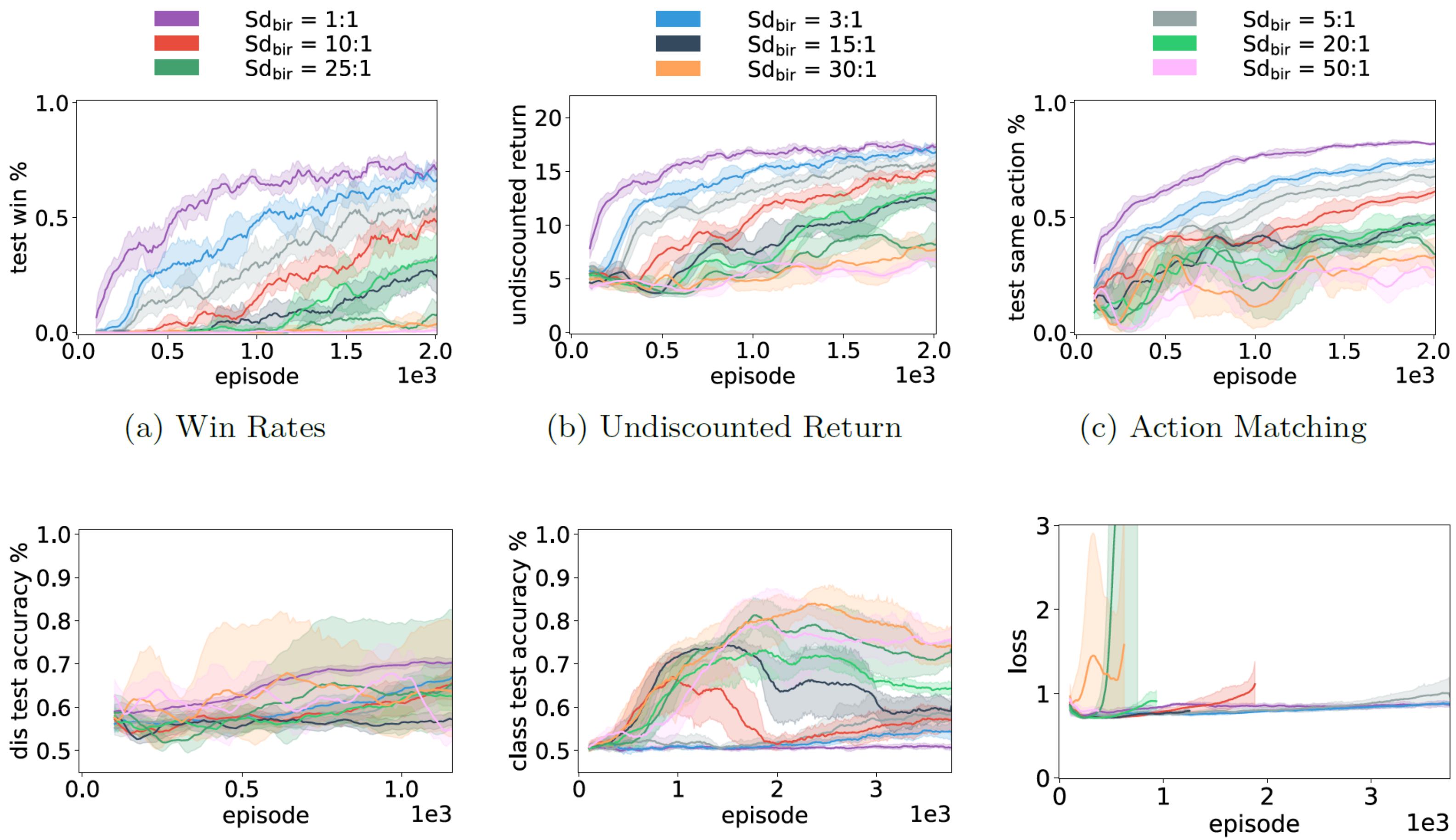
5 Marines vs 6 Marines - Threshold



(a) Win Rates (b) Undiscounted Return (c) Action Matching

(d) Discriminator Accuracy (e) Classifier Accuracy (f) ISAAC Loss

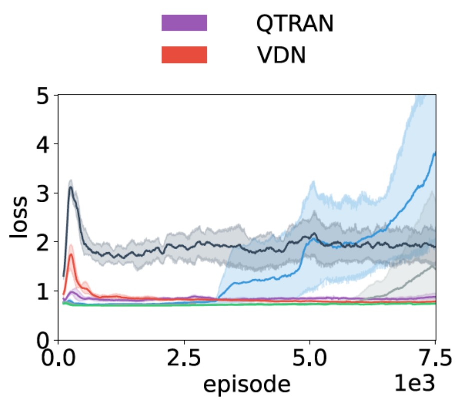
5 Marines vs 6 Marines - Schedules



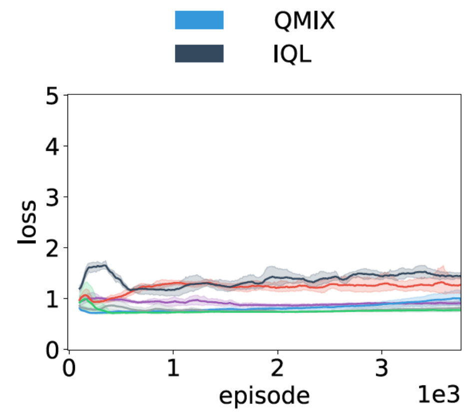
(a) Win Rates (b) Undiscounted Return (c) Action Matching

(d) Discriminator Accuracy (e) Classifier Accuracy (f) ISAAC Loss

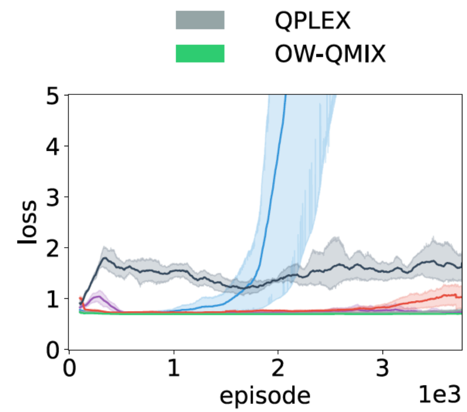
5 Marines vs 6 Marines - Extra



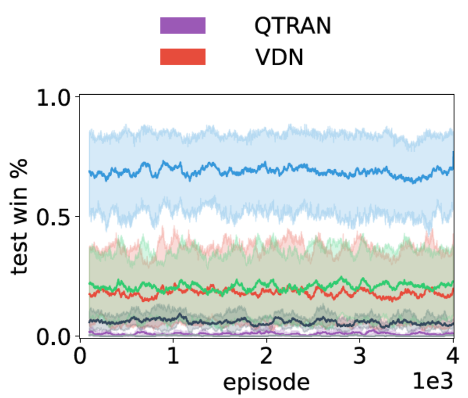
(a) 3s vs 3z ISAAC Loss



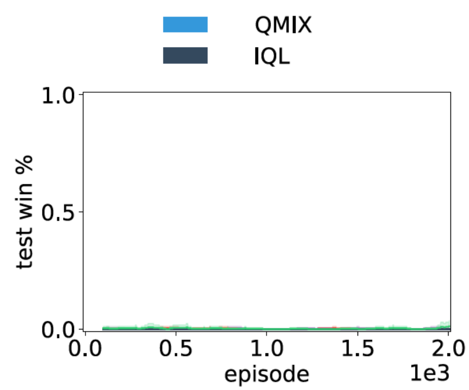
(b) 5m vs 6m ISAAC Loss



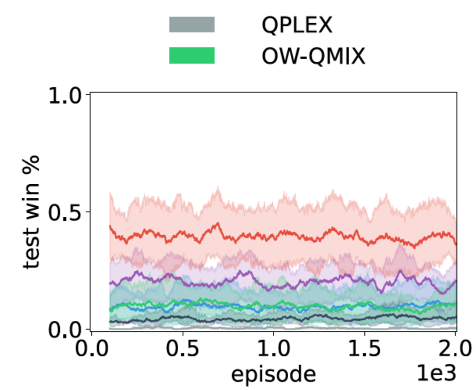
(c) MMM ISAAC Loss



(a) 3s vs 3z Naive Invader Training Test Win Rates

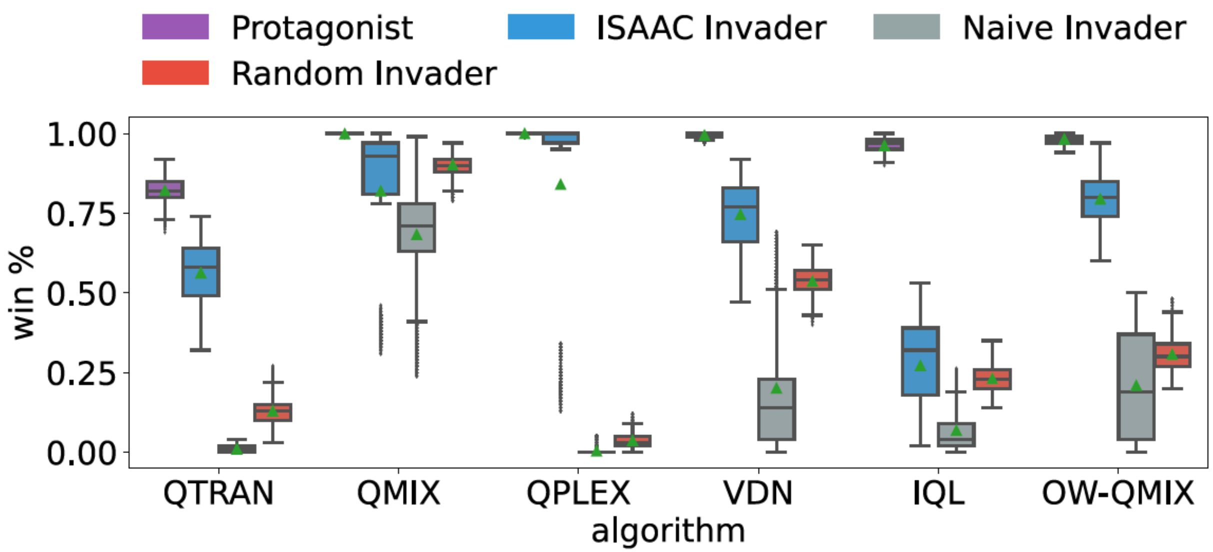


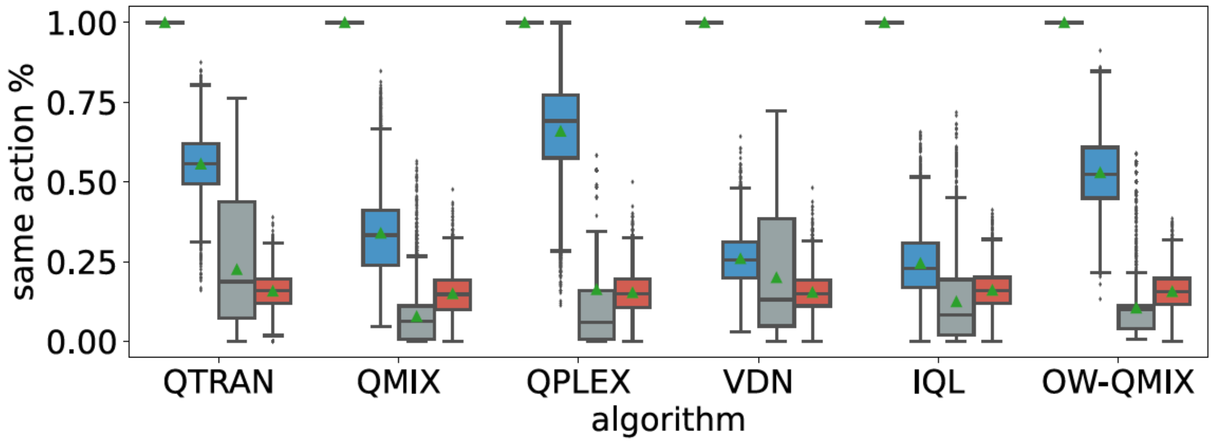
(b) 5m vs 6m Naive Invader Training Test Win Rates



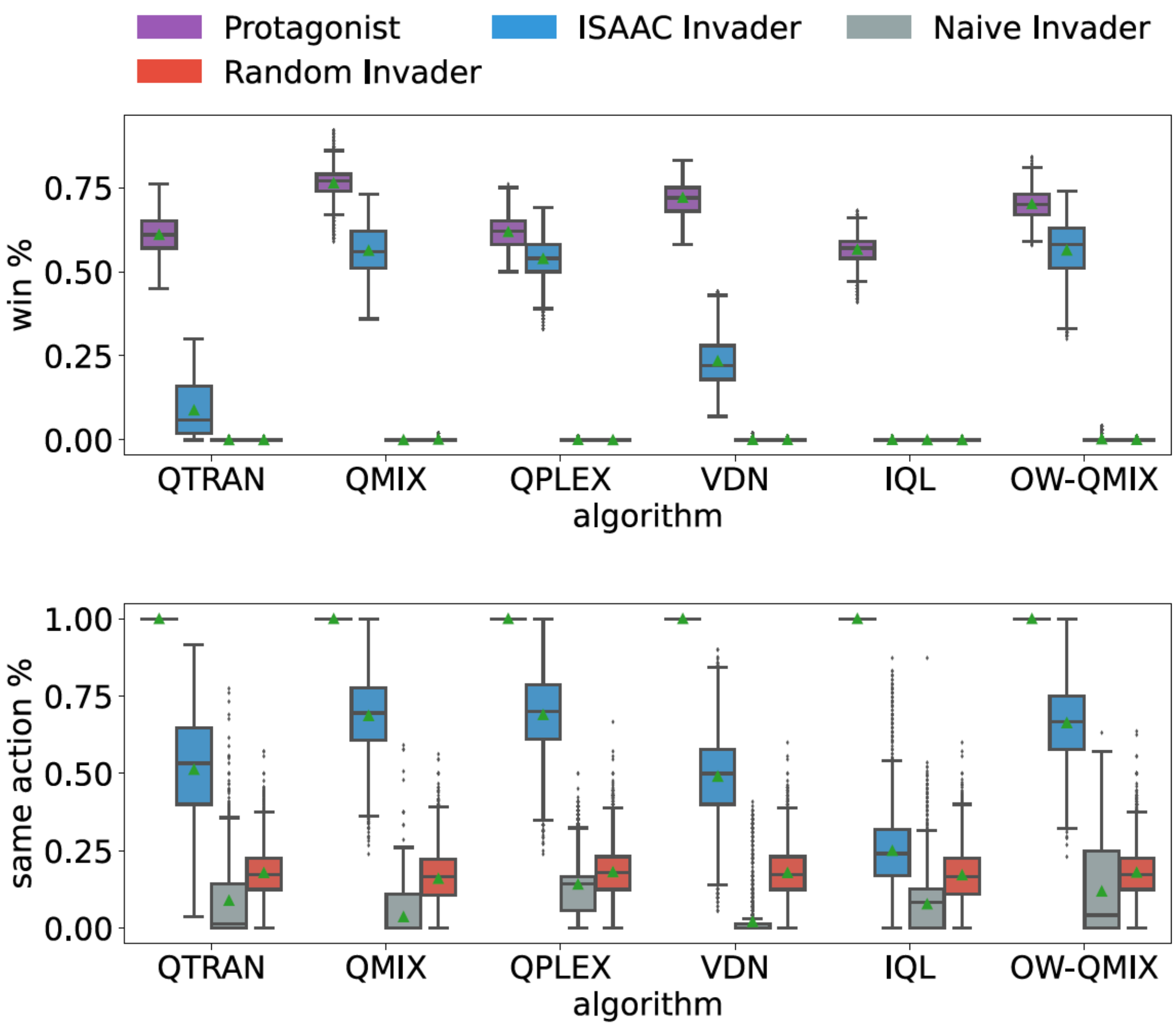
(c) MMM Naive Invader Training Test Win Rates

3 Stalkers vs 3 Zealots - Eval. Box

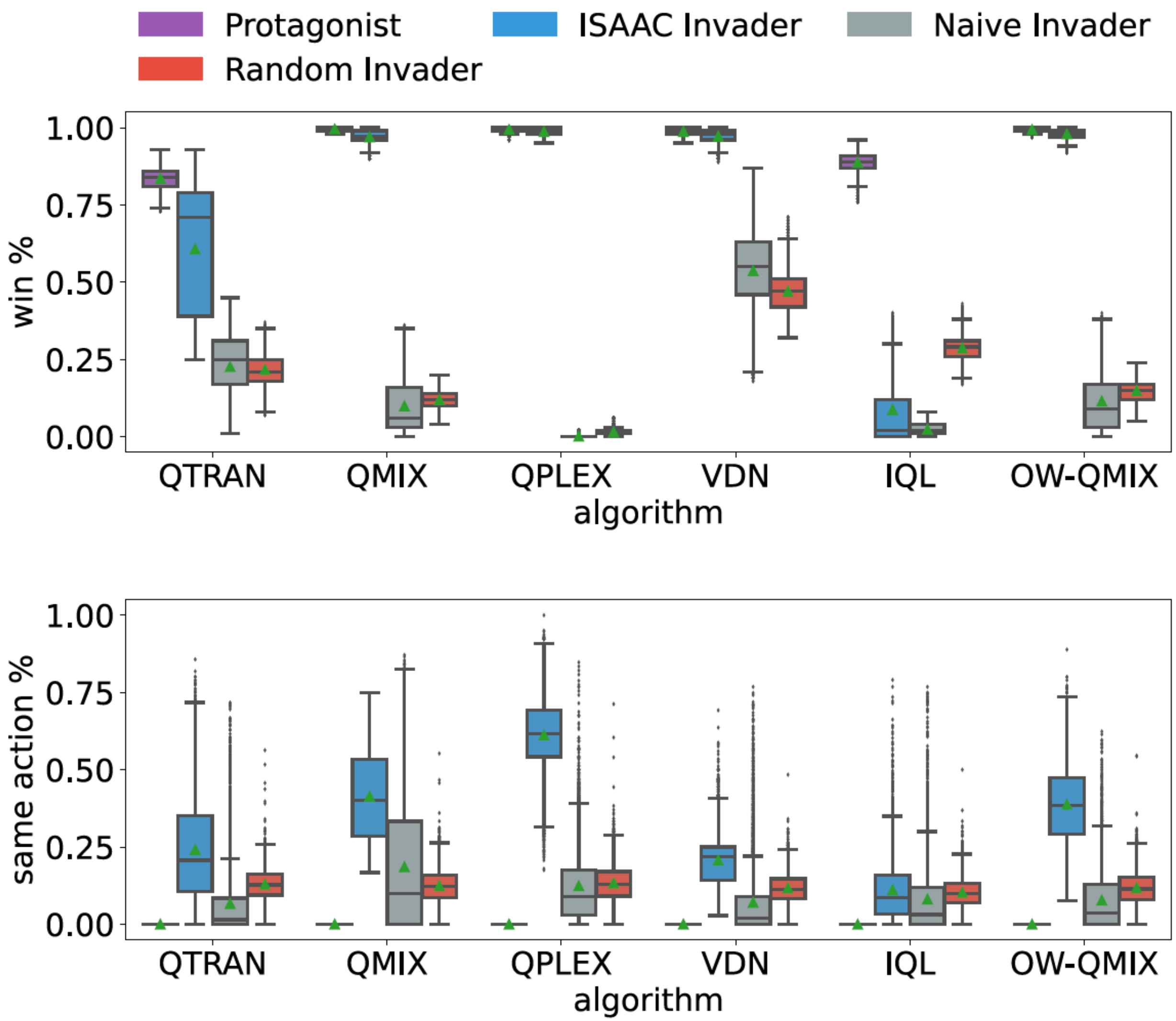




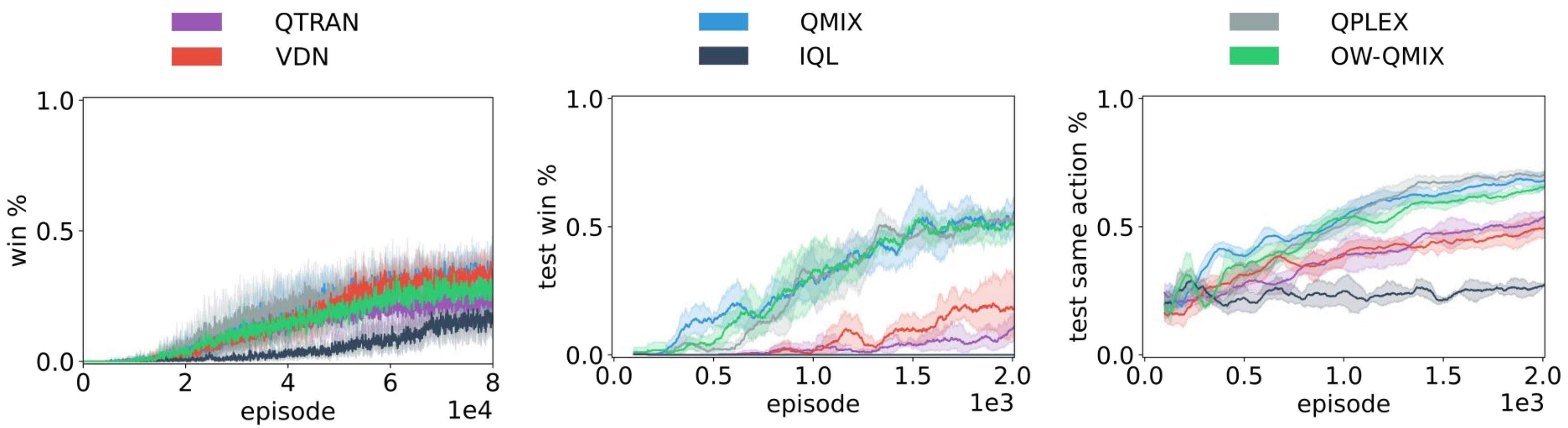
5 Marines vs 6 Marines - Eval. Box



1 Medivac, 2 Marauders & 7 Marines - Eval. Box



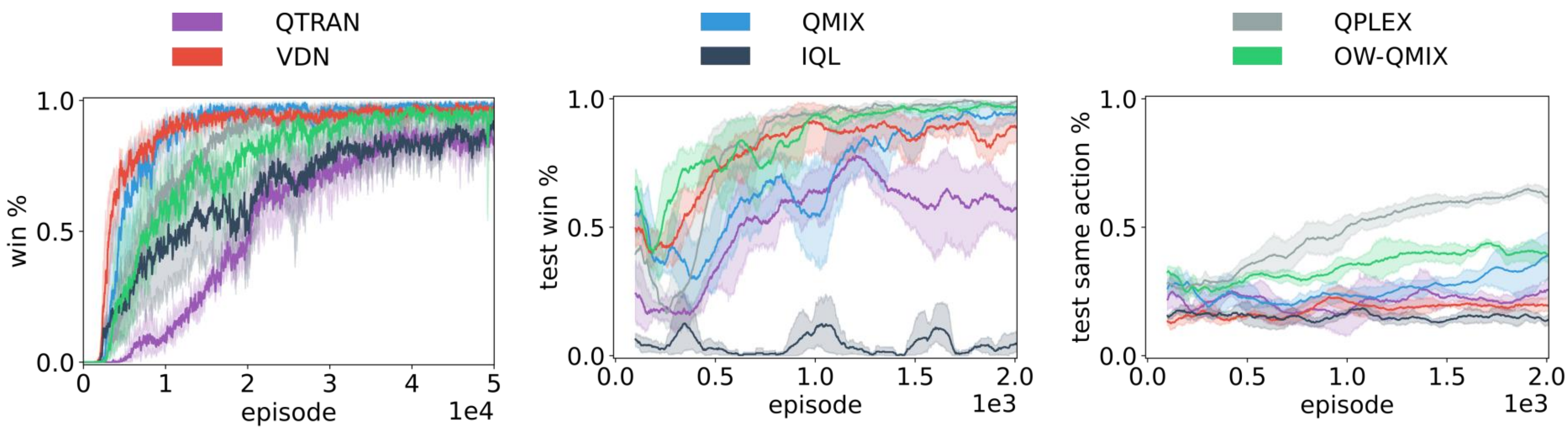
5 Marines vs 6 Marines - Training



Protagonist Training

ISAAC Invader Training

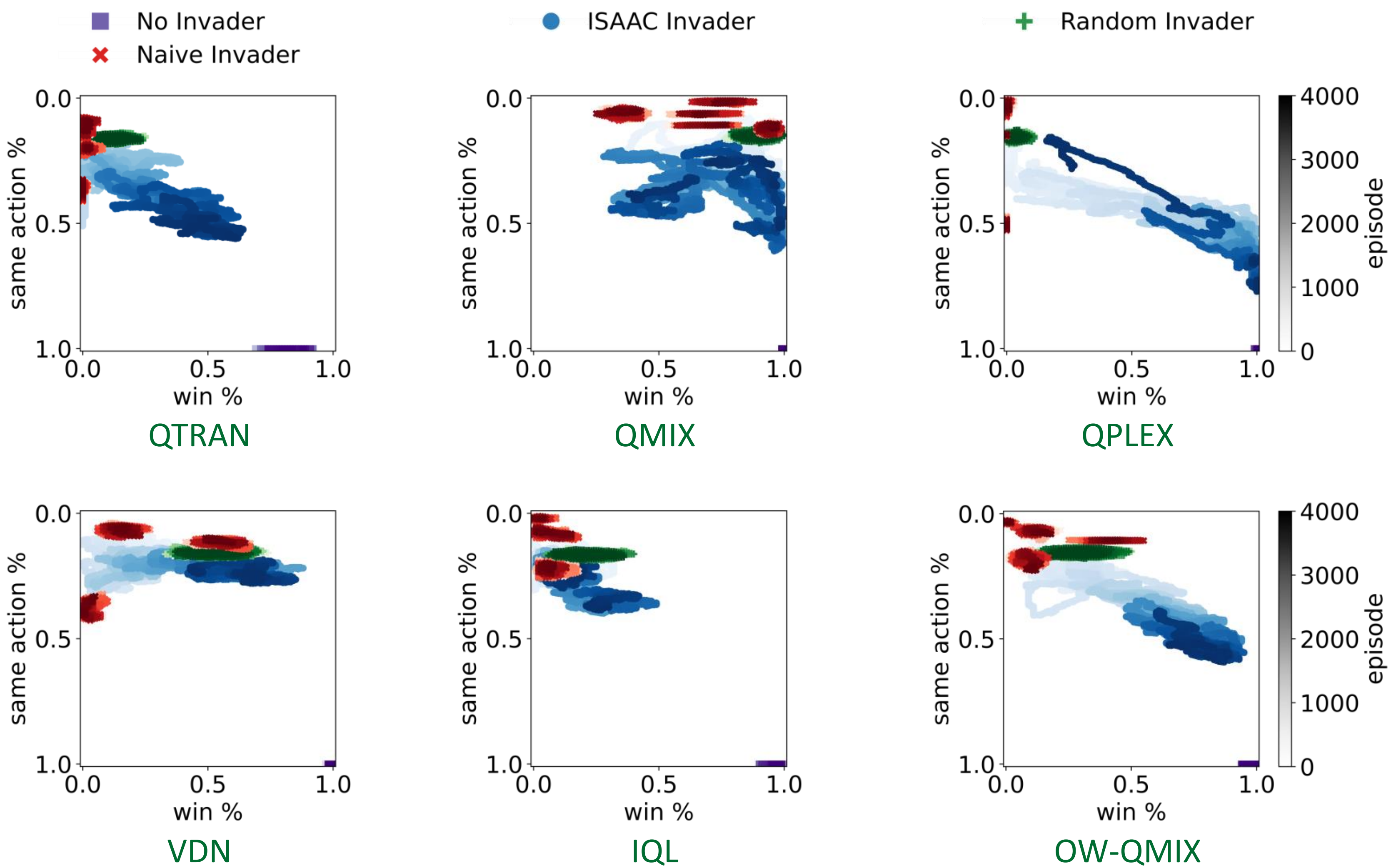
1 Medivac, 2 Marauders & 7 Marines - Training



Protagonist Training

ISAAC Invader Training

3 Stalkers vs 3 Zealots - Trade-off



5 Marines vs 6 Marines - Evaluation

5m_vs_6m	No Substitution	ISAAC Invader	Naive Invader	Random Invader
QTRAN	0.61 ± 0.01	0.09 ± 0.01	0.00 ± 0.00	0.00 ± 0.00
QMIX	0.76 ± 0.01	0.56 ± 0.01	0.00 ± 0.00	0.00 ± 0.00
QPLEX	0.62 ± 0.01	0.54 ± 0.01	0.00 ± 0.00	0.00 ± 0.00
VDN	0.72 ± 0.01	0.23 ± 0.01	0.00 ± 0.00	0.00 ± 0.00
IQL	0.57 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
OW-QMIX	0.70 ± 0.01	0.57 ± 0.01	0.00 ± 0.00	0.00 ± 0.00

Win rates

5m_vs_6m	No Substitution	ISAAC Invader	Naive Invader	Random Invader
QTRAN	1.00 ± 0.00	0.51 ± 0.00	0.09 ± 0.00	0.18 ± 0.00
QMIX	1.00 ± 0.00	0.69 ± 0.00	0.04 ± 0.00	0.16 ± 0.00
QPLEX	1.00 ± 0.00	0.69 ± 0.00	0.14 ± 0.00	0.18 ± 0.00
VDN	1.00 ± 0.00	0.49 ± 0.00	0.02 ± 0.00	0.18 ± 0.00
IQL	1.00 ± 0.00	0.25 ± 0.00	0.08 ± 0.00	0.17 ± 0.00
OW-QMIX	1.00 ± 0.00	0.66 ± 0.00	0.12 ± 0.00	0.18 ± 0.00

Action Matching

1 Medivac, 2 Marauders & 7 Marines - Evaluation

MMM	No Substitution	ISAAC Invader	Naive Invader	Random Invader
QTRAN	0.83 ± 0.01	0.61 ± 0.01	0.23 ± 0.01	0.21 ± 0.01
QMIX	1.00 ± 0.00	0.97 ± 0.00	0.10 ± 0.01	0.12 ± 0.01
QPLEX	0.99 ± 0.00	0.99 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
VDN	0.99 ± 0.00	0.97 ± 0.00	0.54 ± 0.01	0.47 ± 0.01
IQL	0.89 ± 0.01	0.09 ± 0.01	0.02 ± 0.00	0.29 ± 0.01
OW-QMIX	0.99 ± 0.00	0.98 ± 0.00	0.12 ± 0.01	0.15 ± 0.01

Win rates

MMM	No Substitution	ISAAC Invader	Naive Invader	Random Invader
QTRAN	1.00 ± 0.00	0.24 ± 0.00	0.07 ± 0.00	0.13 ± 0.00
QMIX	1.00 ± 0.00	0.41 ± 0.00	0.19 ± 0.01	0.12 ± 0.00
QPLEX	1.00 ± 0.00	0.61 ± 0.00	0.12 ± 0.00	0.13 ± 0.00
VDN	1.00 ± 0.00	0.21 ± 0.00	0.07 ± 0.00	0.12 ± 0.00
IQL	1.00 ± 0.00	0.11 ± 0.00	0.08 ± 0.00	0.10 ± 0.00
OW-QMIX	1.00 ± 0.00	0.39 ± 0.00	0.08 ± 0.00	0.12 ± 0.00

Action Matching

Code

Algorithm 1 Infiltrating Stealth Agent Attack Controller (ISAAC)

Initialize Protagonist and ISAAC Invader replay memories D_{pro}, D_{atk}
Load trained protagonist parameters θ into θ_{pro} for π_{pro}
Initialize θ_{atk} for π_{atk} with random parameters θ_{rand}
Initialize θ_{dis} and θ_{class} with random parameters θ_{rand}
Initialize Schedules Sd_{epi}, Sd_{bir}, Sd_f
Create attack set $\mathcal{S}^a_{comb} = \mathcal{S}^a_{pro} \cup \mathcal{S}^a_{atk}$

for epi = 1 to M **do**
    $\mathcal{S}^a, D \leftarrow$ select $\mathcal{S}^a_{comb}, D_{atk}$ or $\mathcal{S}^a_{pro}, D_{pro}$ according to $Sd_{epi}(epi)$
    Run one episode using $\mathcal{S}^a$
    Store all global history experience tuples $\boldsymbol{e}^{\boldsymbol{\tau}}_t = \langle s_t, z_t, \boldsymbol{\tau}_t, \boldsymbol{u}_t, r_t, s_{t+1}, z_{t+1}, \boldsymbol{\tau}_{t+1} \rangle$ in D
    Sample random minibatch B_{pro}, B_{atk} from D_{pro}, D_{atk}
    $\mathcal{L}_{imit}, \mathcal{L}_{dis}, \mathcal{L}_{class} \leftarrow$ Set to 0

    if $Sd_{bir}(epi)$ **then**
        $z, \tau \leftarrow$ take observations and observation histories from B_{atk}
        $u_{pro}, u_{atk} \leftarrow$ action selection according to $\pi_{pro}(z), \pi_{atk}(z)$
        $v_{pro}, v_{atk} \leftarrow$ discriminator input according to $(u_{pro}, \tau), (u_{atk}, \tau)$
        $\mathcal{L}_{dis} \leftarrow H_p(v_{pro}, v_{atk})$ w.r.t. Eq. 5.1
        $\mathcal{L}_{imit} \leftarrow H_{bs,p}(v_{pro}, v_{atk})$ w.r.t. Eq. 5.1 and followed by Eq. 5.2
    end if

    $r_F \leftarrow$ update all rewards r in B_{atk} using classifier w.r.t. Eq. 5.3
    $\mathcal{L}_{ant} \leftarrow \mathcal{L}(\theta_{atk})$ w.r.t. Eq. 2.1 with $Q^{\boldsymbol{\pi}}(\boldsymbol{\tau}_t, \boldsymbol{u}_t; \theta) = -Q^{\boldsymbol{\pi}}(\boldsymbol{\tau}_t, \boldsymbol{u}_t; \theta)$, B_{atk} and $r = r_F$

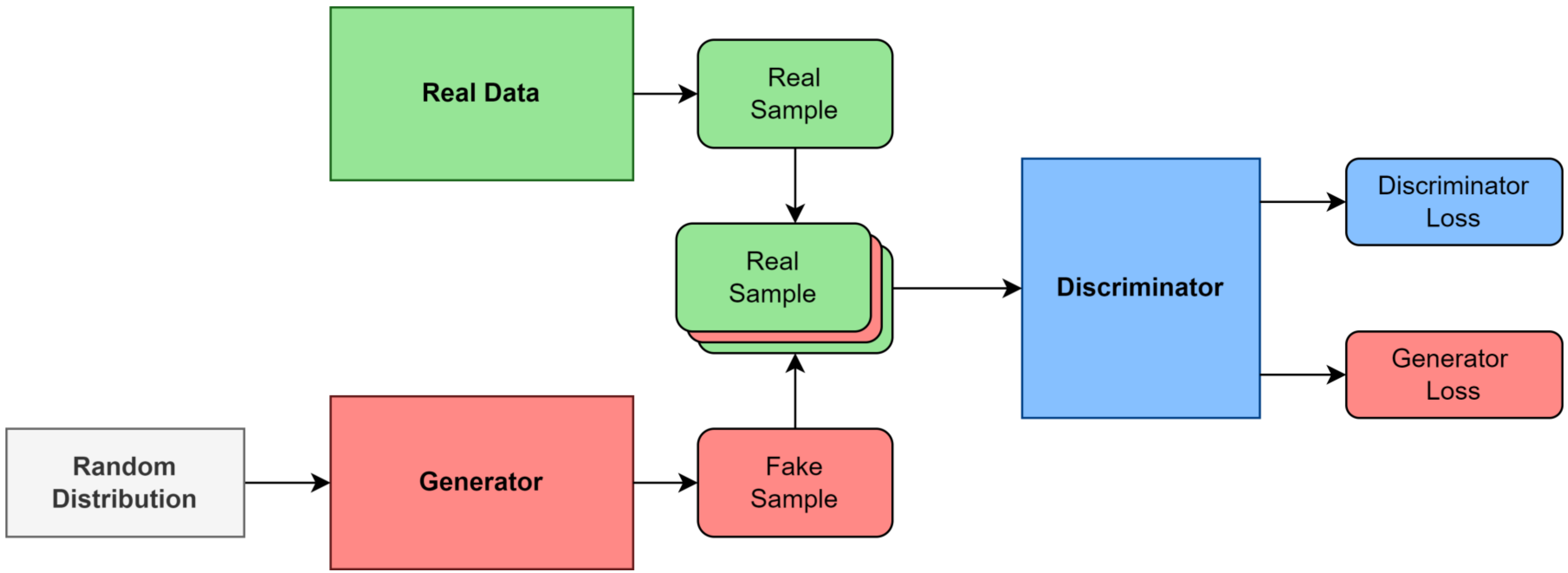
    $\mathcal{L}_{atk} \leftarrow \mathcal{L}_{ant} + \mathcal{L}_{imit}$

    if $Sd_f(epi)$ **then**
        $s'_{pro}, s'_{atk} \leftarrow$ select state features from B_{pro}, B_{atk}
        $\mathcal{L}_{class} \leftarrow H_p(s'_{pro}, s'_{atk})$ w.r.t. Eq. 2.12
    end if

    Update θ_{dis} by minimizing $\mathcal{L}_{dis}$
    Update θ_{class} by minimizing $\mathcal{L}_{class}$
    Update θ_{atk} by minimizing $\mathcal{L}_{atk}$

end for

Generative Adversarial Networks



$$\min_G \max_D B(G,D) = \min_G \max_D \mathbb{E}_{x \sim p_x}[\log D(x)] + \mathbb{E}_{z \sim p_d}[\log(1 - D(G(z)))]$$

Source: Goodfellow et al. (2014) **Generative Adversarial Nets**

Bibliography

[ABC+20] Charles Audet, Jean Bigeon, Dominique Cartier, Sébastien Le Digabel, and Ludovic Salomon. Performance indicators in multiobjective optimization. *European Journal of Operational Research*, 292, 11 2020.

[ACB17] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017.

[AFLZ22] Abolfazl Ahmadi, Mohammad Mahdi Forootan, Iman Larki, and Rahim Zahedi. Machine learning and deep learning in energy systems: A review. *Sustainability*, 14, 09 2022.

[AK19] Nahid Anwar and Susmita Kar. Review paper on various software testing techniques strategies. *Global Journal of Computer Science and Technology*, pages 43–49, 05 2019.

[And86] Charles W. Anderson. Learning and problem-solving with multilayer connectionist systems (adaptive, strategy learning, neural networks, reinforcement learning). 1986.

[And17] H. Anderson. Evading machine learning malware detection. 2017.

[ASM+11] Prince Agarwal, Manjari Sahai, Vaibhav Mishra, Monark Bag, and Vrijendra Singh. A review of multi-criteria decision making techniques for supplier evaluation and selection. *International Journal of Industrial Engineering Computations*, 2:801–810, 10 2011.

[Bar18] Samuel A. Barnett. Convergence problems with generative adversarial networks (gans). *CoRR*, abs/1806.11382, 2018.

[BEH+20] Yoram Bachrach, Richard Everett, Edward Hughes, Angeliki Lazaridou, Joel Z. Leibo, Marc Lanctot, Michael Johanson, Wojciech M. Czarnecki, and Thore Graepel. Negotiating team formation using deep reinforcement learning. *CoRR*, abs/2010.10380, 2020.

[Bel57] Richard Bellman. A markovian decision process. *Indiana Univ. Math. J.*, 6:679–684, 1957.

[BH22] Hojat Behrooz and Yeganeh M. Hayeri. Machine learning applications in surface transportation systems: A literature review. *Applied Sciences*, 12(18), 2022.

[BMR+20] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. *CoRR*, abs/2005.14165, 2020.

[BSGK22] Junyoung Byun, Kyujin Shim, Hyojun Go, and Changick Kim. Hidden conditional adversarial attacks. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1306–1310, 2022.

[BZI13] Daniel S. Bernstein, Shlomo Zilberstein, and Neil Immerman. The complexity of decentralized control of markov decision processes. *CoRR*, abs/1301.3836, 2013.

[CDS19] Aidan Clark, Jeff Donahue, and Karen Simonyan. Efficient video generation on complex datasets. *CoRR*, abs/1907.06571, 2019.

[CEW11] Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. Computational aspects of cooperative game theory. volume 5, 10 2011.

[Che21] Haiyang Chen. Challenges and corresponding solutions of generative adversarial networks (gans): A survey study. *Journal of Physics: Conference Series*, 1827(1):012066, mar 2021.

[CLX+19] Tong Chen, Jiqiang Liu, Yingxiao Xiang, Wenjia Niu, Endong Tong, and Zhen Han. Adversarial attack and defense in reinforcement learning-from ai security view. *Cybersecurity*, 2, 12 2019.

[CPLK20] Mingyu Cai, Hao Peng, Zhijun Li, and Zhen Kan. Learning-based probabilistic ltl motion planning with environment and motion uncertainties. *IEEE Transactions on Automatic Control*, PP, 07 2020.

[CvMBB14] KyungHyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259, 2014.

[CW16] Nicholas Carlini and David A. Wagner. Towards evaluating the robustness of neural networks. *CoRR*, abs/1608.04644, 2016.

[DMW+20] Ranjie Duan, Xingjun Ma, Yisen Wang, James Bailey, A. Kai Qin, and Yun Yang. Adversarial camouflage: Hiding physical-world attacks with natural styles. *CoRR*, abs/2003.08757, 2020.

[GAZ17] Amirata Ghorbani, Abubakar Abid, and James Zou. Interpretation of neural networks is fragile, 2017.

[GEB15] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. *CoRR*, abs/1508.06576, 2015.

[GLZ+20] Ruize Gao, Feng Liu, Jingfeng Zhang, Bo Han, Tongliang Liu, Gang Niu, and Masashi Sugiyama. Maximum mean discrepancy is aware of adversarial attacks. *CoRR*, abs/2010.11415, 2020.

[Goo17] Ian J. Goodfellow. NIPS 2016 tutorial: Generative adversarial networks. *CoRR*, abs/1701.00160, 2017.

[GPAM+14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[GRBN09] Jinwei Gu, Ravi Ramamoorthi, Peter Belhumeur, and Shree Nayar. Removing image artifacts due to dirty camera lenses and thin occluders. *ACM Trans. Graph.*, 28(5):1–10, dec 2009.

[GSS14] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples, 2014.

[Gun18] Nyoman Gunantara. A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, 5(1):1502242, 2018.

[GWH19] Jan Philip Göpfert, Heiko Wersing, and Barbara Hammer. Adversarial attacks hidden in plain sight. *CoRR*, abs/1902.09286, 2019.

[GY18] Abhishek Gupta and Zhaoyuan Yang. Adversarial reinforcement learning for observer design in autonomous systems under cyber attacks. *CoRR*, abs/1809.06784, 2018.

[HBD21] Keyang He, Bikramjit Banerjee, and Prashant Doshi. Cooperative-competitive reinforcement learning with history-dependent rewards. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '21, page 602–610, Richland, SC, 2021. International Foundation for Autonomous Agents and Multiagent Systems.

[HCC+21] Rongjie Huang, Chenye Cui, Feiyang Chen, Yi Ren, Jinglin Liu, Zhou Zhao, Baoxing Huai, and Zhefeng Wang. Singgan: Generative adversarial network for high-fidelity singing voice generation, 2021.

[HE16] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. *CoRR*, abs/1606.03476, 2016.

[HEKK18] Sara Hooker, Dumitru Erhan, Pieter-Jan Kindermans, and Been Kim. Evaluating feature importance estimates. *CoRR*, abs/1806.10758, 2018.

[HPG+17] Sandy H. Huang, Nicolas Papernot, Ian J. Goodfellow, Yan Duan, and Pieter Abbeel. Adversarial attacks on neural network policies. *CoRR*, abs/1702.02284, 2017.

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

[HS15] Matthew J. Hausknecht and Peter Stone. Deep recurrent q-learning for partially observable mdps. *CoRR*, abs/1507.06527, 2015.

[HZ19] Yunhan Huang and Quanyan Zhu. Deceptive reinforcement learning under adversarial manipulations on cost signals. *CoRR*, abs/1906.10571, 2019.

[HZB+19] Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial examples. *CoRR*, abs/1907.07174, 2019.

[HZRS15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *CoRR*, abs/1502.01852, 2015.

[IRM+19] Mohsen Imani, Mohammad Saidur Rahman, Nate Mathews, Aneesh Yogesh Joshi, and Matthew Wright. Adv-dwf: Defending against deep-learning-based website fingerprinting attacks with adversarial traces. *CoRR*, abs/1902.06626, 2019.

Bibliography

[JAAA16] Abid Jamil, Muhammad Arif, Normi Abubakar, and Akhlaq Ahmad. Software testing techniques: A literature review. pages 177–182, 11 2016.

[JEP+21] John M. Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Andy Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David A. Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596:583 – 589, 2021.

[JSJ94] Tommi Jaakkola, Satinder P. Singh, and Michael I. Jordan. Reinforcement learning algorithm for partially observable markov decision problems. In *Proceedings of the 7th International Conference on Neural Information Processing Systems*, NIPS'94, page 345–352, Cambridge, MA, USA, 1994. MIT Press.

[KdW06] I. Y. Kim and Olivier L. de Weck. Adaptive weighted sum method for multi-objective optimization: a new method for pareto front generation. *Structural and Multidisciplinary Optimization*, 31:105–116, 2006.

[KGB16] Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio. Adversarial machine learning at scale. *CoRR*, abs/1611.01236, 2016.

[KGT+0] Sai Rahul Kaminwar, Jann Goschenhofer, Janek Thomas, Ingo Thon, and Bernd Bischl. Structured verification of machine learning models in industrial settings. *Big Data*, 0(0):null, 0. PMID: 34978896.

[Kis19] Michael Kissner. Hacking neural networks: A short introduction. *CoRR*, abs/1911.07658, 2019.

[KL51] Solomon Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[KLA18] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *CoRR*, abs/1812.04948, 2018.

[KNLH19] Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey E. Hinton. Similarity of neural network representations revisited. *CoRR*, abs/1905.00414, 2019.

[KP19] Stepan Komkov and Aleksandr Petiushko. Advhat: Real-world adversarial attack on arcface face ID system. *CoRR*, abs/1908.08705, 2019.

[KS17] Jernej Kos and Dawn Song. Delving into adversarial attacks on deep policies, 2017.

[LHL+17] Yen-Chen Lin, Zhang-Wei Hong, Yuan-Hong Liao, Meng-Li Shih, Ming-Yu Liu, and Min Sun. Tactics of adversarial attack on deep reinforcement learning agents. *CoRR*, abs/1703.06748, 2017.

[Lin91] Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Trans. Inf. Theory*, 37:145–151, 1991.

[Lit94] Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In William W. Cohen and Haym Hirsh, editors, *Machine Learning Proceedings 1994*, pages 157–163. Morgan Kaufmann, San Francisco (CA), 1994.

[LLM+22] Cameron Lischke, Tongtong Liu, Joe McCalmon, Md Asifur Rahman, Talal Halabi, and Sarra Alqahtani. Lstm-based anomalous behavior detection in multi-agent reinforcement learning. In *2022 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 16–21, 2022.

[LMK+17] Jan Leike, Miljan Martic, Victoria Krakovna, Pedro A. Ortega, Tom Everitt, Andrew Lefrancq, Laurent Orseau, and Shane Legg. AI safety gridworlds. *CoRR*, abs/1711.09883, 2017.

[LSC21] Ken Ming Lee, Sriram Ganapathi Subramanian, and Mark Crowley. Investigation of independent reinforcement learning algorithms in multi-agent environments. *CoRR*, abs/2111.01100, 2021.

[LZL+17] Joel Z. Leibo, Vinícius Flores Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. *CoRR*, abs/1702.03037, 2017.

[McC95] R. Andrew McCallum. Instance-based utile distinctions for reinforcement learning with hidden state. In Armand Prieditis and Stuart J. Russell, editors, *Proceedings of the Twelfth International Conference on Machine Learning (ICML 1995)*, pages 387–395, San Francisco, CA, USA, 1995. Morgan Kauffman.

[Mes18] Lars M. Mescheder. On the convergence properties of GAN training. *CoRR*, abs/1801.04406, 2018.

[MFF15] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: a simple and accurate method to fool deep neural networks. *CoRR*, abs/1511.04599, 2015.

[MKS+13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

[MKS+15] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015.

[MMS+17] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks, 2017.

[MMS+19] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *CoRR*, abs/1908.09635, 2019.

[MO14] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.

[MRSW19] Anuj Mahajan, Tabish Rashid, Mikayel Samvelyan, and Shimon Whiteson. MAVEN: multi-agent variational exploration. *CoRR*, abs/1910.07483, 2019.

[Nak05] Hirotaka Nakayama. Multi-objective optimization and its engineering applications. In Jürgen Branke, Kalyanmoy Deb, Kaisa Miettinen, and Ralph E. Steuer, editors, *Practical Approaches to Multi-Objective Optimization, 7.-12. November 2004*, volume 04461 of *Dagstuhl Seminar Proceedings*. IBFI, Schloss Dagstuhl, Germany, 2005.

[Net05] Goncalo Neto. From single-agent to multi-agent reinforcement learning: Foundational concepts and methods. 01 2005.

[NH19] Atsuhiro Noguchi and Tatsuya Harada. Image generation from small datasets via batch statistics adaptation. *CoRR*, abs/1904.01774, 2019.

[NIGM18] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *CoRR*, abs/1811.03378, 2018.

[NMSE20] Elior Nehemya, Yael Mathov, Asaf Shabtai, and Yuval Elovici. Taking over the stock market: Adversarial perturbations against algorithmic traders, 2020.

[Oli12] Frans A. Oliehoek. *Decentralized POMDPs*, pages 471–503. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[OSV11] Frans A. Oliehoek, Matthijs T. J. Spaan, and Nikos Vlassis. Optimal and approximate q-value functions for decentralized pomdps. *CoRR*, abs/1111.0062, 2011.

[PDSG17] Lerrel Pinto, James Davidson, Rahul Sukthankar, and Abhinav Gupta. Robust adversarial reinforcement learning. *CoRR*, abs/1703.02702, 2017.

[PGS+20] Thomy Phan, Thomas Gabor, Andreas Sedlmeier, Fabian Ritz, Bernhard Kempter, Cornel Klein, Horst Sauer, Reiner Schmid, Jan Wieghardt, Marc Zeller, and Claudia Linnhoff-Popien. Learning and testing resilience in cooperative multi-agent systems. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '20, page 1055–1063, Richland, SC, 2020. International Foundation for Autonomous Agents and Multiagent Systems.

[PP19] Roberta Pappadà and Francesco Pauli. Discrimination in machine learning algorithms. 01 2019.

[Pro07] Emily Pronin. Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, 11:37–43, 2007.

[PTL+17] Anay Pattanaik, Zhenyi Tang, Shuijing Liu, Gautham Bommannan, and Girish Chowdhary. Robust deep reinforcement learning with adversarial attacks. *CoRR*, abs/1712.03632, 2017.

[QYZ+21] You Qiaoben, Chengyang Ying, Xinning Zhou, Hang Su, Jun Zhu, and Bo Zhang. Understanding adversarial attacks on observations in deep reinforcement learning. *CoRR*, abs/2106.15860, 2021.

Bibliography

[RBL+21] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. *CoRR*, abs/2112.10752, 2021.

[RDN+22] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents, 2022.

[RFPW20] Tabish Rashid, Gregory Farquhar, Bei Peng, and Shimon Whiteson. Weighted QMIX: expanding monotonic value function factorisation. *CoRR*, abs/2006.10800, 2020.

[RRD+20] Amin Rakhsha, Goran Radanovic, Rati Devidze, Xiaojin Zhu, and Adish Singla. Policy teaching via environment poisoning: Training-time adversarial attacks against reinforcement learning. *CoRR*, abs/2003.12909, 2020.

[RSdW+18] Tabish Rashid, Mikayel Samvelyan, Christian Schröder de Witt, Gregory Farquhar, Jakob N. Foerster, and Shimon Whiteson. QMIX: monotonic value function factorisation for deep multi-agent reinforcement learning. *CoRR*, abs/1803.11485, 2018.

[RSL18] Aditi Raghunathan, Jacob Steinhardt, and Percy Liang. Certified defenses against adversarial examples. *CoRR*, abs/1801.09344, 2018.

[Rub99] Reuven Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability*, 1(2):127–190, 1999.

[Rud16] Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016.

[SB98] R.S. Sutton and A.G. Barto. *Reinforcement learning: an introduction*. MIT Press, Cambridge, MA., 1998.

[SBBR16] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 1528–1540, New York, NY, USA, 2016. Association for Computing Machinery.

[SCHL21] Lucas Schott, Manon Césaire, Hatem Hajri, and Sylvain Lamprier. Improving robustness of deep reinforcement learning agents: Environment attacks based on critic networks. *CoRR*, abs/2104.03154, 2021.

[SGZ+16] Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *CoRR*, abs/1606.03498, 2016.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[SKK+19] Kyunghwan Son, Daewoo Kim, Wan Ju Kang, David Earl Hostallero, and Yung Yi. Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning, 2019.

[SLG+17] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinícius Flores Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z. Leibo, Karl Tuyls, and Thore Graepel. Value-decomposition networks for cooperative multi-agent learning. *CoRR*, abs/1706.05296, 2017.

[SMRL59] Alvin Scodel, J. Sayer Minas, Philburn Ratoosh, and Milton E. Lipetz. Some descriptive aspects of two-person non-zero-sum games. *Journal of Conflict Resolution*, 3:114 – 119, 1959.

[SRdW+19] Mikayel Samvelyan, Tabish Rashid, Christian Schröder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philip H. S. Torr, Jakob N. Foerster, and Shimon Whiteson. The starcraft multi-agent challenge. *CoRR*, abs/1902.04043, 2019.

[Sut88] Richard Sutton. Learning to predict by the method of temporal differences. *Machine Learning*, 3:9–44, 08 1988.

[SVJ19] S. Reema Sree, S.B. Vyshnavi, and N. Jayapandian. Real-world application of machine learning and deep learning. In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 1069–1073, 2019.

[SVS17] Jiawei Su, Danilo Vasconcellos Vargas, and Kouichi Sakurai. One pixel attack for fooling deep neural networks. *CoRR*, abs/1710.08864, 2017.

[SWW+19] Shawn Shan, Emily Willson, Bolun Wang, Bo Li, Haitao Zheng, and Ben Y. Zhao. Gotta catch 'em all: Using concealed trapdoors to detect adversarial attacks on neural networks. *CoRR*, abs/1904.08554, 2019.

[SZS+13] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks, 2013.

[SZZ18] Kun Shao, Yuanheng Zhu, and Dongbin Zhao. Starcraft micromanagement with reinforcement learning and curriculum transfer learning. *CoRR*, abs/1804.00810, 2018.

[Tan93] Ming Tan. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Machine Learning Proceedings 1993*, pages 330–337. Morgan Kaufmann, San Francisco (CA), 1993.

[TTV18] Hoang Thanh-Tung, Truyen Tran, and Svetha Venkatesh. On catastrophic forgetting and mode collapse in generative adversarial networks. *CoRR*, abs/1807.04015, 2018.

[TWW+21] James Tu, Tsun-Hsuan Wang, Jingkang Wang, Sivabalan Manivasagam, Mengye Ren, and Raquel Urtasun. Adversarial attacks on multi-agent communication. *CoRR*, abs/2101.06560, 2021.

[UO12] Alberto Uriarte and Santiago Ontanon. Kiting in rts games using influence maps. pages 31–36, 01 2012.

[VAAMS22] Kevin Voogd, Jean Pierre Allamaa, Javier Alonso-Mora, and Tong Duy Son. Reinforcement learning from simulation to real world autonomous driving using digital twin, 2022.

[VBC+19] Oriol Vinyals, Igor Babuschkin, Wojciech Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John Agapiou, Max Jaderberg, and David Silver. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575, 11 2019.

[VH13] Mark Velasquez and Patrick Hester. An analysis of multi-criteria decision making methods. *International Journal of Operations Research*, 10:56–66, 05 2013.

[VRQI10] Sebastian Varges, Giuseppe Riccardi, Silvia Quarteroni, and Alexei Ivanov. The exploration/exploitation trade-off in reinforcement learning for dialogue management. pages 479 – 484, 01 2010.

[WD92] Christopher Watkins and Peter Dayan. Technical note: Q-learning. *Machine Learning*, 8:279–292, 05 1992.

[WFL+22] Kuan-Chieh Wang, Yan Fu, Ke Li, Ashish Khisti, Richard S. Zemel, and Alireza Makhzani. Variational model inversion attacks. *CoRR*, abs/2201.10787, 2022.

[WLS+22] Zhongshu Wang, Lingzhi Li, Zhen Shen, Li Shen, and Liefeng Bo. 4k-nerf: High fidelity neural radiance fields at ultra high resolutions, 2022.

[WPKD12] J.B. Williams, D. Popp, K.A. Kobak, and M.J. Detke. P-640 - the power of expectation bias. *European Psychiatry*, 27:1, 2012. Abstracts of the 20th European Congress of Psychiatry.

[WRL+20] Jianhao Wang, Zhizhou Ren, Terry Liu, Yang Yu, and Chongjie Zhang. QPLEX: duplex dueling multi-agent q-learning. *CoRR*, abs/2008.01062, 2020.

[WZH17] Huikai Wu, Junge Zhang, and Kaiqi Huang. MSC: A dataset for macro-management in starcraft II. *CoRR*, abs/1710.03131, 2017.

[XEZJ22] Zikang Xiong, Joe Eappen, He Zhu, and Suresh Jagannathan. Defending observation attacks in deep reinforcement learning via detection and denoising, 2022.

[XI15] Shuo Xiong and Hiroyuki Iida. Attractiveness of real time strategy games. *2014 2nd International Conference on Systems and Informatics, ICSAI 2014*, 12:271–276, 01 2015.

[XLWW21] Wenju Xu, Chengjiang Long, Ruisheng Wang, and Guanghui Wang. DRB-GAN: A dynamic resblock generative adversarial network for artistic style transfer. *CoRR*, abs/2108.07379, 2021.

[XWF+18] Zheng Xu, Michael J. Wilber, Chen Fang, Aaron Hertzmann, and Hailin Jin. Beyond textures: Learning from multi-domain artistic images for arbitrary style transfer. *CoRR*, abs/1805.09987, 2018.

[YHZ+17] Xiaoyong Yuan, Pan He, Qile Zhu, Rajendra Rana Bhat, and Xiaolin Li. Adversarial examples: Attacks and defenses for deep learning. *CoRR*, abs/1712.07107, 2017.

Bibliography

[YJvdS19] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[YLD71] Yv Haimes Yv, Leon S. Lasdon, and Dang Da. On a bicriterion formation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man, and Cybernetics*, pages 296–297, 1971.

[ZCBH21] Huan Zhang, Hongge Chen, Duane S. Boning, and Cho-Jui Hsieh. Robust reinforcement learning on state observations with learned optimal adversary. *CoRR*, abs/2101.08452, 2021.

[ZDS17] Zhengli Zhao, Dheeru Dua, and Sameer Singh. Generating natural adversarial examples. *CoRR*, abs/1710.11342, 2017.

[ZHML19] Jie M. Zhang, Mark Harman, Lei Ma, and Yang Liu. Machine learning testing: Survey, landscapes and horizons. *CoRR*, abs/1906.10742, 2019.

[ZML16] Junbo Jake Zhao, Michaël Mathieu, and Yann LeCun. Energy-based generative adversarial network. *CoRR*, abs/1609.03126, 2016.

[ZWZ22] Jie Zhu, Fengge Wu, and Junsuo Zhao. An overview of the action space for deep reinforcement learning. In *2021 4th International Conference on Algorithms, Computing and Artificial Intelligence*, ACAI'21, New York, NY, USA, 2022. Association for Computing Machinery.