AlphaStar: Grandmaster level in StarCraft II Explained

Yekun Chai

Institute of Automation, Chinese Academy of Sciences

yekun.chai@ia.ac.cn

November 12, 2019

Overview

StarCraft II: What and Why

- Real-time strategy game: gather resources, build technology, defeat opponent
- Complexity: among video games, considered to be at the peak of human ability
- Canonical: played by millions, esport endured 20 years of active human play
- Research[1]: hundreds of submissions over 12 years of competition

[1] credit to David Silver

The challenge of StarCraft II

StarCraft represents a major challenge for real-world AI[2]:

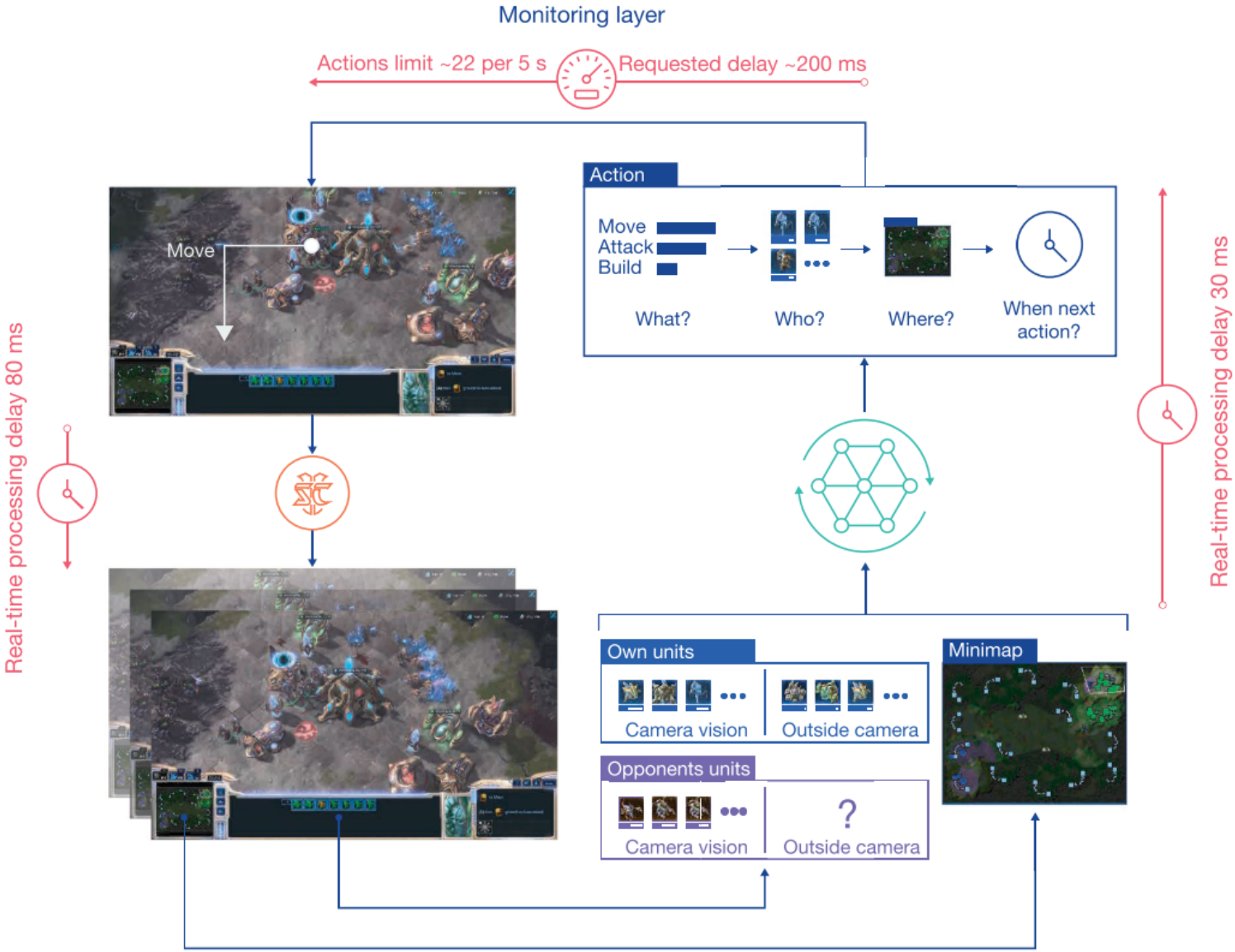
- **Partial observability** - only see information in the camera view
- **Imperfect information** - only see opponent units within range of own units
- **Large action space** - simultaneous control of hundreds of units
- **Strategy cycles** - counter-strategies discovered by pro players over 20 years

[2] credit to David Silver

How does AlphaStar work?

1. At step t, the agent receives an observation o_t (imperfect).
2. For each action a_t with $\approx 10^{26}$ possible choices:
   - what action type
   - who to issue that action to
   - where to target
   - when to observe and act next
3. Limit reaction time and action rates (APM 22 per 5 secs)

How does AlphaStar work?



How does it train?

Supervised learning (main contribution!)

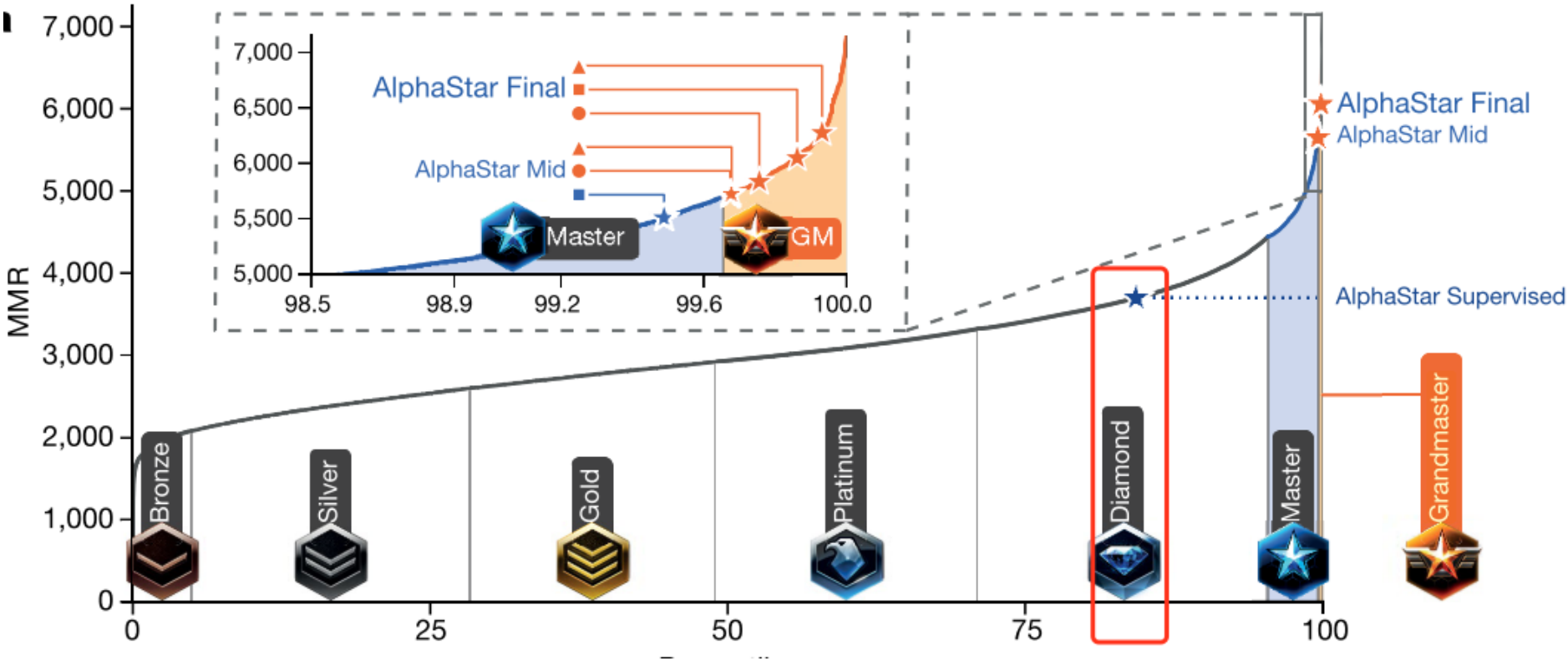
- AlphaStar is initially trained with **supervised learning** (SL) with anonymous game replays from human experts
- This offers a good initialization for neural networks
- This initial agent beat the built-in "Elite" level AI ($\approx$ human golden level)

Reinforcement learning

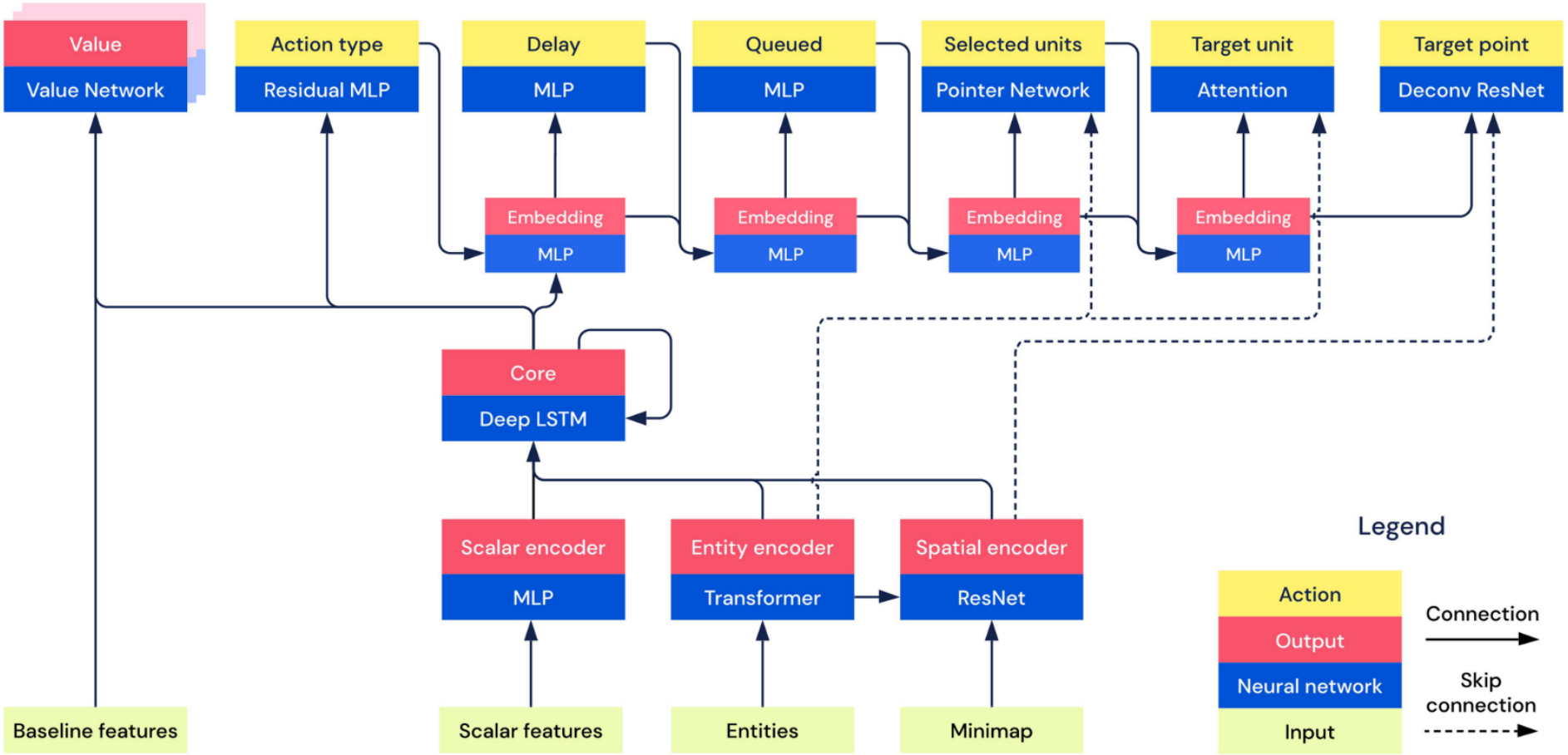
- $\theta_0 \leftarrow \theta_{\mathrm{SL}}$
- Model $\pi_\theta(a_t|s_t, z) = \mathbb{P}[a_t|s_t, z]$, where z summarizes the strategies sampled from a human data
- Seed a multi-agent RL with a continuous league
- Adopt *population-based training* (PBT) and *multi-agent* RL
- The league training can be treated as a bootstrapping DAgger process

Contributions of SL and RL

Pure supervised training could put the agent at the Diamond level!

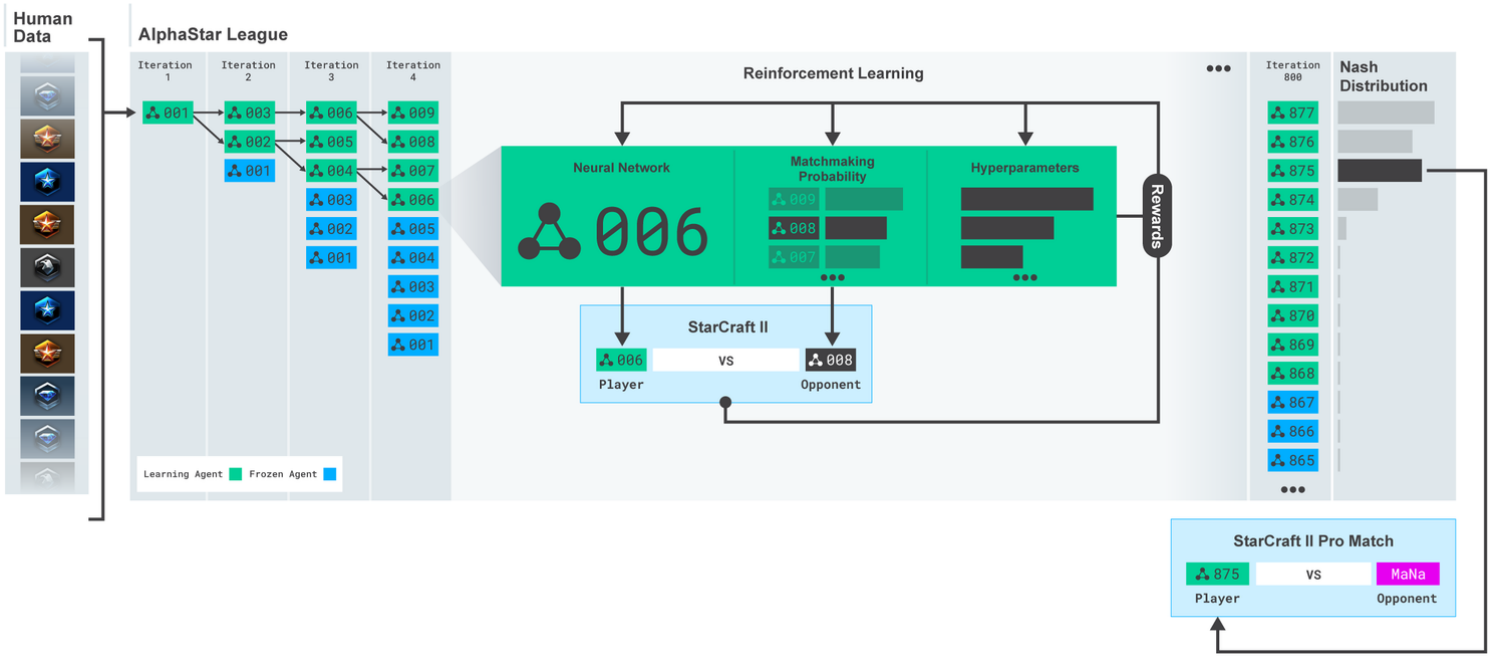


AlphaStar Architecture



AlphaStar league training

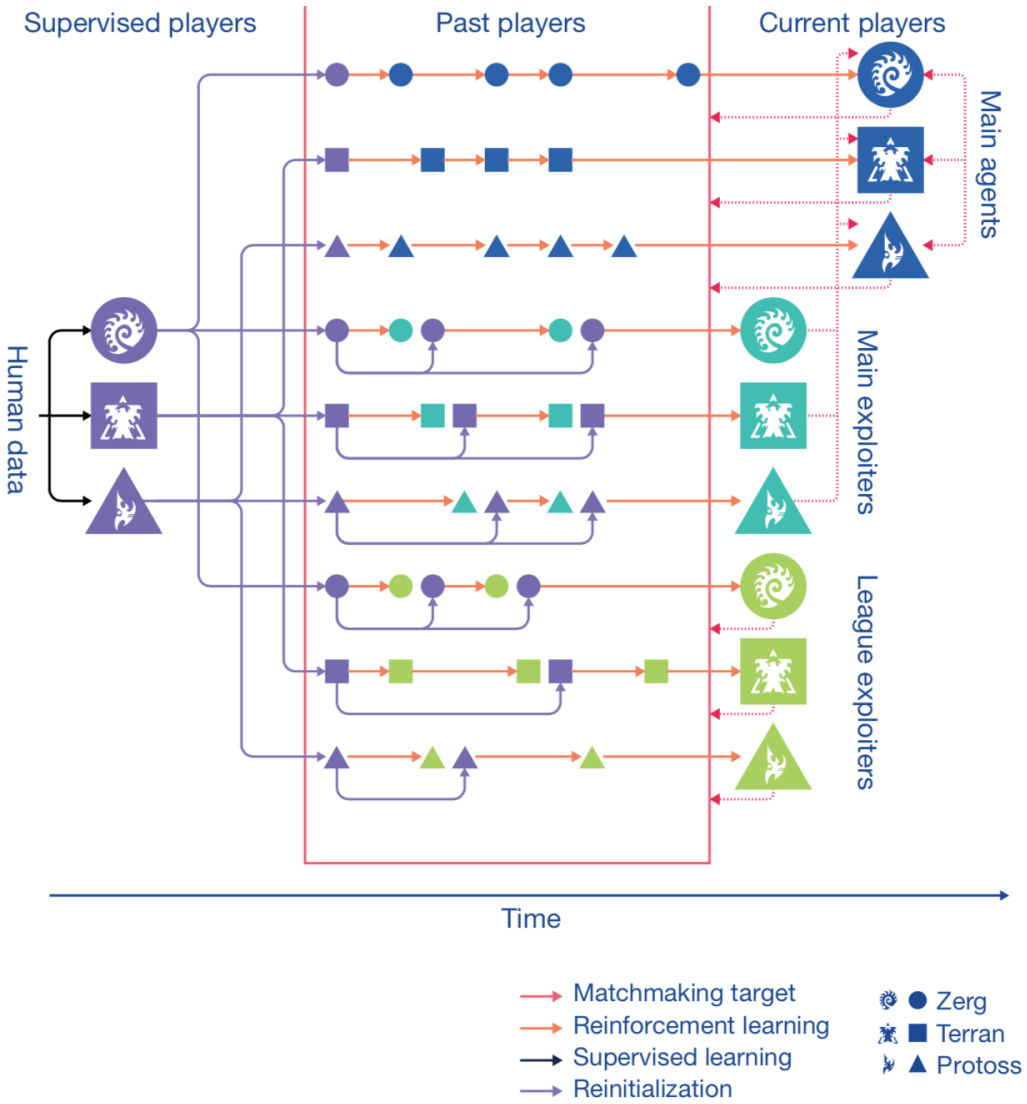
- Form a continuous league wherein agents competing with each other
- New agents were dynamically supplemented to the league, by branching from existing competitors; each agent then learns from games against other competitors.
- Each agent would play against the strongest strategies and does not forget how to defeat the earlier version of agents.



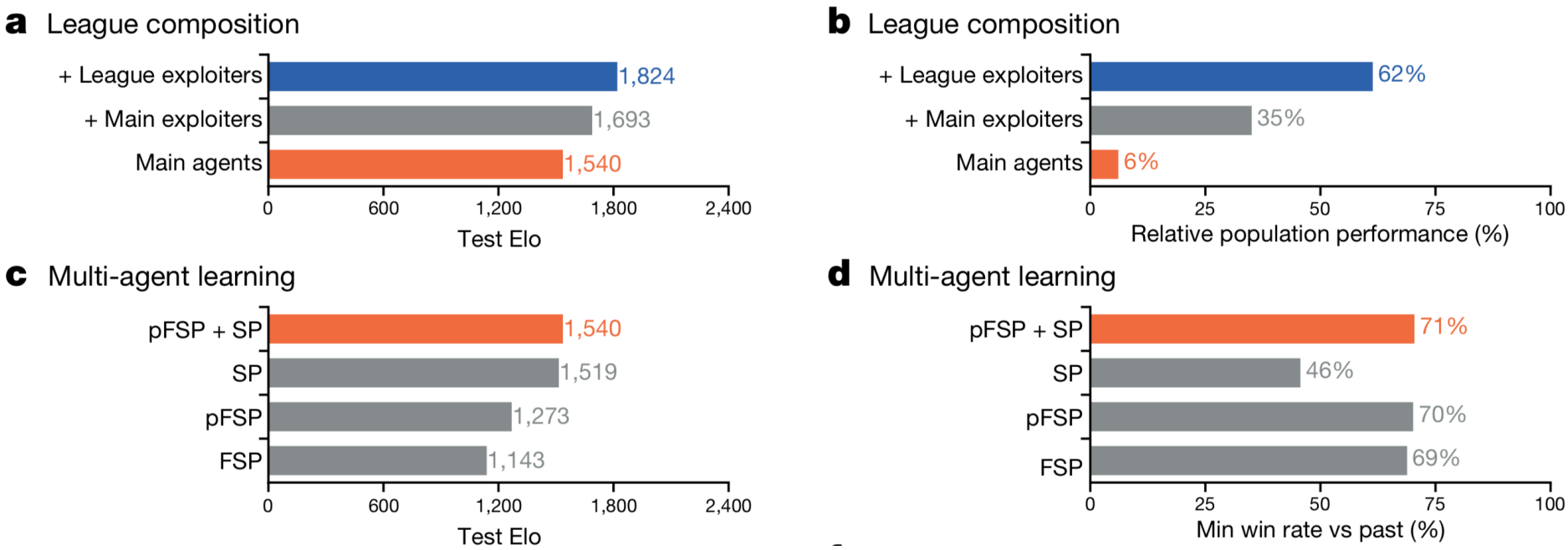
League training explained

1. Put initialized agents into the league, and divide them into *main exploiters*, *league exploiters* and *league exploiters*.
   - *main agents* - against past players and themselves
   - *main exploiters* - against main agents
   - *league exploiters* - against all past players
2. When adding a player to the league, reset main exploiters and league exploiters to supervised agents.
3. Matchmaking strategies: prioritised Fictious Self-Play (pFSP)

League training

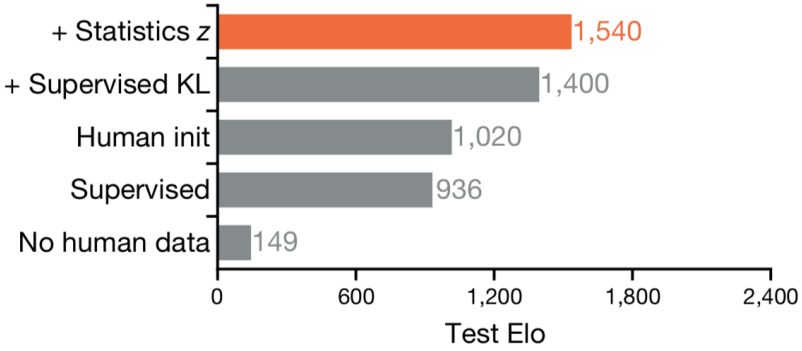


Key components of AlphaStar

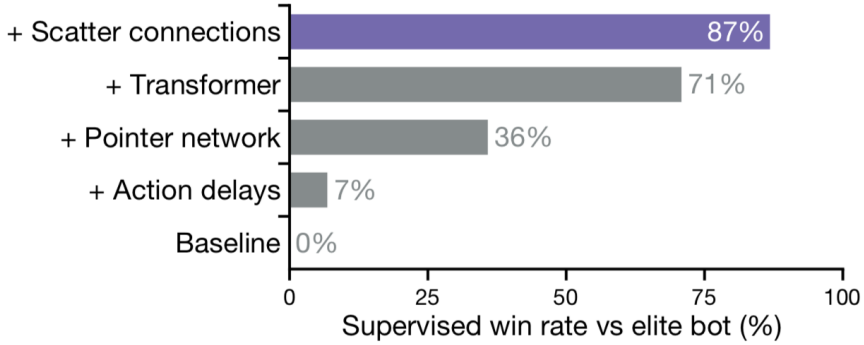


Key components of AlphaStar

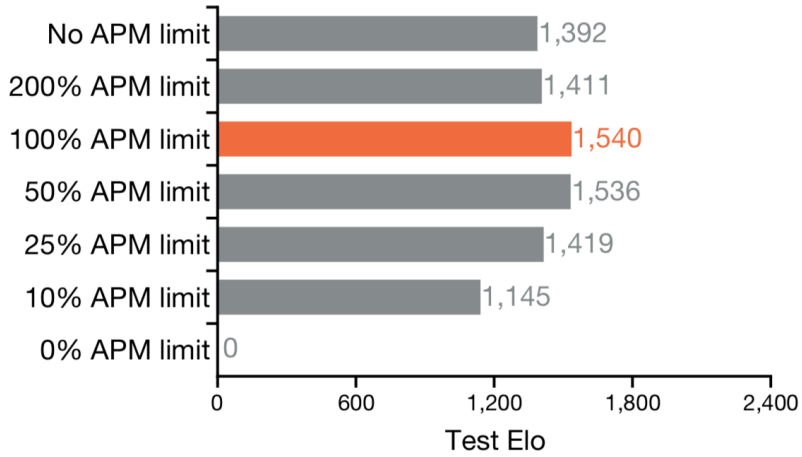
e Human data usage



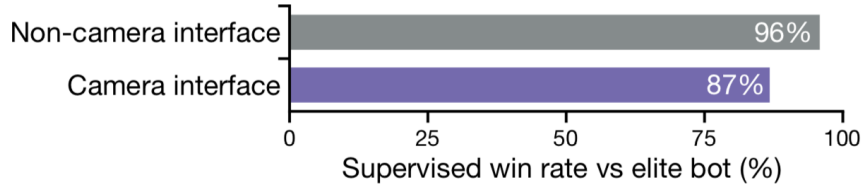
f Architectures



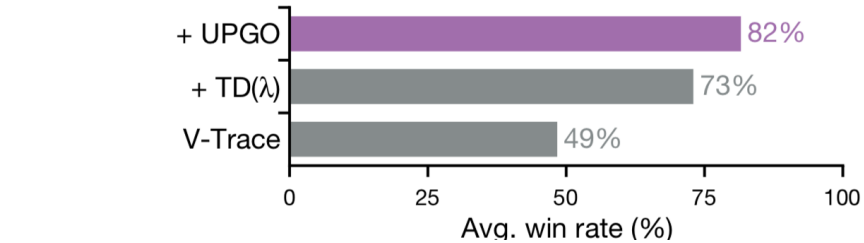
g APM limits



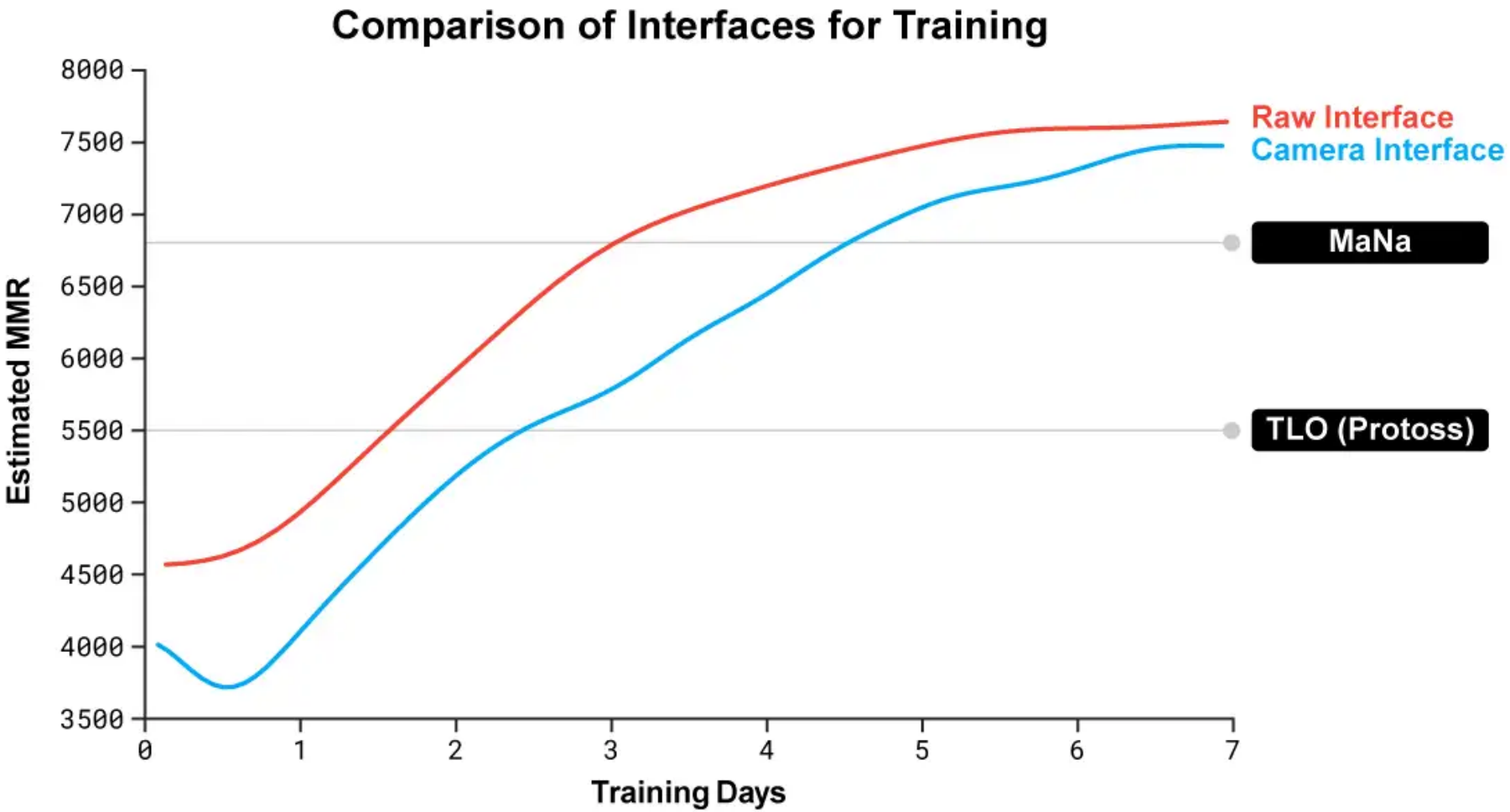
h Interface



i Off-policy learning



Feature engineering also matters!



Thoughts of diversity and exploration

- Naive exploration in micro-tactics could lead to a huge waste of computation. AlphaStar adopts z statistics to maintain the diversity.
- Grounding on this, add constraints like $\mathbb{KL}(\theta|\theta_{SL})$.
- In terms of the building order, use Edit Distance / Hamming Distance to serve as pseudo-rewards to avoid naive exploration

References

1. Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., ... Oh, J. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature, 1-5.
2. Vinyals, O., Ewalds, T., Bartunov, S., Georgiev, P., Vezhnevets, A. S., Yeo, M., ... Quan, J. (2017). Starcraft II: A new challenge for reinforcement learning. arXiv preprint arXiv:1708.04782
3. https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii