Supplementary Material of 'Grid-Wise Control for Multi-Agent Reinforcement Learning in Video Game AI'

A.1 Explanation for Terms and Supplementary Results

Term	Description
Ground unit	units on the ground that may collide with each other when moving
Air unit	air units will not collide with each other
Melee unit	ground unit that can only attack enemies standing close to it
Range unit	unit that can attack enemies in distance
Zealot	a melee unit of Protoss
Immortal	a range unit of Protoss
Baneling	ground unit that explodes to produce range damage to enemy ground units
Zergling	ground melee unit that attacks enemy ground units
Roach	ground range unit that attacks enemy ground units
Hydralisk	ground range unit that can attack both ground and air units
Mutalisk	air range unit that can attack both ground and air units

Table 5. Explanations of terms in StarCraft II appeared in Section 5.1.

Table 6. Average cross combat winning rate of the 4 trained policies of each method on 5I. The bold text indicates the strongest policy which can achieve above 50% winning rate against all the other 3 policies.

		Rand	AN	HR	SP
IAC	Rand	–	1.00	1.00	0.02
	AN	0.00	–	0.41	0.00
	HR	0.00	0.59	–	0.00
	SP	0.98	1.00	1.00	–
IQL	Rand	–	0.30	0.61	0.50
	AN	0.70	–	0.66	0.51
	HR	0.39	0.34	–	0.51
	SP	0.50	0.49	0.49	–
Central-V	Rand	–	0.23	0.22	0.23
	AN	0.77	–	0.45	0.95
	HR	0.78	0.55	–	0.67
	SP	0.77	0.05	0.33	–
CommNet	Rand	–	0.42	0.39	0.02
	AN	0.58	–	0.65	0.21
	HR	0.61	0.35	–	0.15
	SP	0.98	0.79	0.85	–
GridNet	Rand	–	0.40	0.37	0.24
	AN	0.60	–	0.50	0.12
	HR	0.63	0.50	–	0.11
	SP	0.76	0.88	0.89	–

Table 7. Average cross combat winning rate of the 4 trained policies of each method on 3I2Z. The bold text indicates the strongest policy which can achieve above 50% winning rate against all the other 3 policies.

		Rand	AN	HR	SP
IAC	Rand	–	0.97	1.00	0.24
	AN	0.03	–	0.58	0.00
	HR	0.00	0.42	–	0.00
	SP	0.76	1.00	1.00	–
IQL	Rand	–	0.37	0.90	0.59
	AN	0.63	–	0.71	0.63
	HR	0.10	0.29	–	0.13
	SP	0.41	0.37	0.87	–
Central-V	Rand	–	0.29	0.14	0.62
	AN	0.71	–	0.37	0.73
	HR	0.86	0.63	–	0.73
	SP	0.38	0.27	0.27	–
CommNet	Rand	–	0.39	0.23	0.47
	AN	0.61	–	0.46	0.62
	HR	0.77	0.54	–	0.68
	SP	0.53	0.38	0.32	–
GridNet	Rand	–	0.54	0.52	0.44
	AN	0.46	–	0.57	0.40
	HR	0.48	0.43	–	0.35
	SP	0.56	0.60	0.65	–

A.2 Detailed Experimental Settings

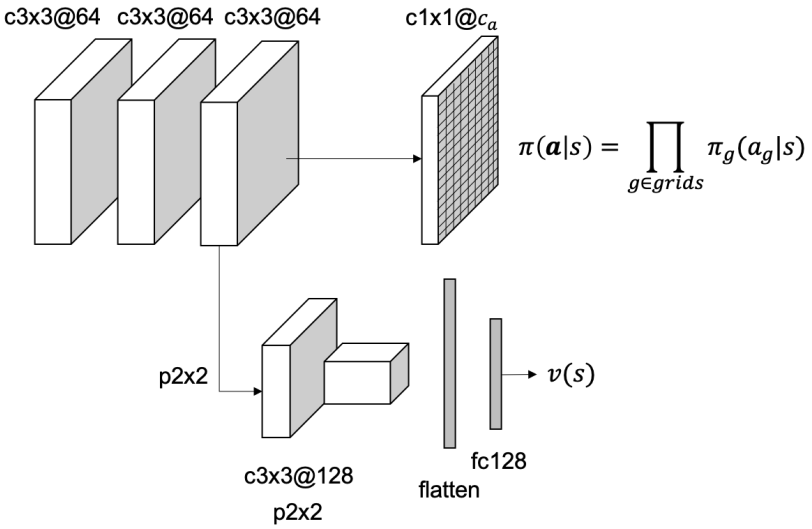
For 5I, the size of the grid feature map is $(8, 8, 6)$ with 6 channels including the health points (HP), shield and cooling down (CD) for both sides. For 3I2Z, the size of the grid feature map is $(8, 8, 16)$ with 16 channels including the HP, shield, CD and damage for both Immortal and Zealot on both sides. We find that on these two scenarios, a map of size 8×8 is sufficient to let GridNet work well. For MAB, the size of the grid feature map is $(16, 16, 18)$ with 18 channels including Mutalisk HP, Hydralisk HP, Roach HP, Zergling HP, Baneling HP, CD, attack range, damage and whether it can attack air units for both sides. The number of units from each unit type will be randomized over episodes in the range $[1, 5]$. On MAB, we use a larger feature map with size 16×16, since there might exist much more agents in this setting.

It is worth mentioning that the Mutalisk is an air unit and it does not collide with other units, so on MAB more than one units could overlap in the same grid. In such a case, we first distinguish the air units and the ground units in channels. That is, we use different channels in the observation and action maps for the ground and air units. Then, for all the air units, if more than one units overlap in a grid, we randomly select one of them to control. We find this trick works well in practice.

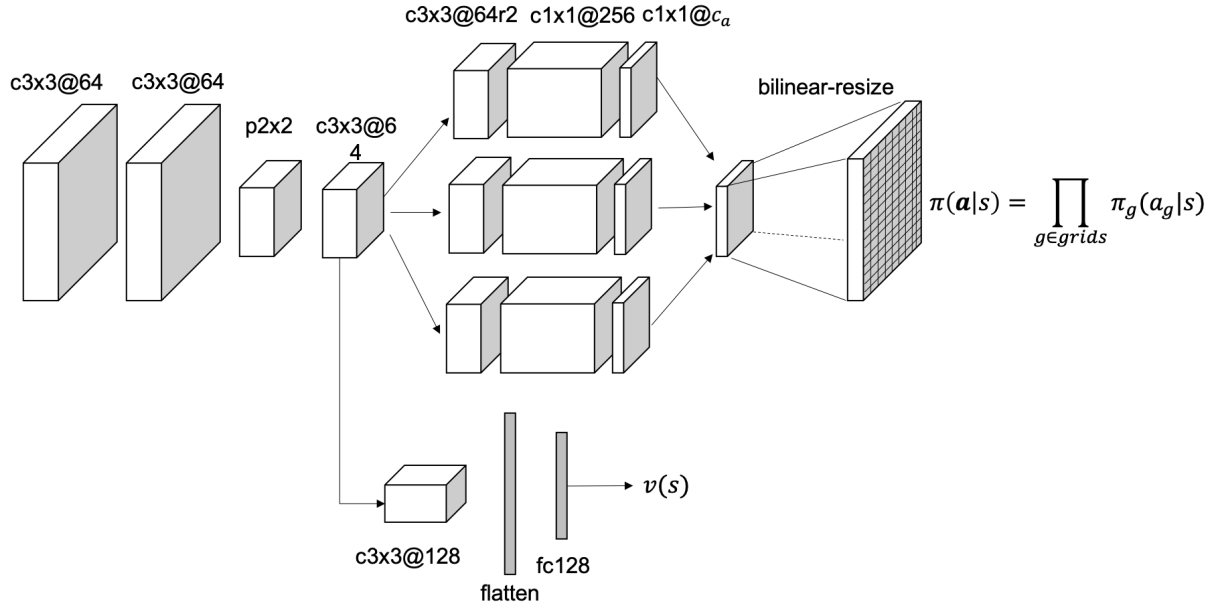
In addition to the grid feature map, the compared methods take as input an additional individual status vector for each agent, including the general information appeared in the channels of the grid feature map plus the coordinates.

A.3 Neural Network Structures

The neural network structures used for GridNet are depicted in Fig. 6. On 5I and 3I2Z, the grid map is small and we propose to use standard convolutional neural networks without reducing the map size. For MAB, we adopt a structure similar to the DeepLabV3 (Chen et al., 2018). All of the other compared methods adopt exactly the same structure as the encoder of GridNet for their grid feature map, and they use additional 2 fc layers with size 64 to process the individual status vector of each agent. Then, the output of the convolutional encodings are flattened into size of 64 and concatenated with the output of the fc layers, and then connect 2-3 fc layers of size 32 to output the policy or value, conditioning on different methods. The CommNet contains a specific communication layer, which is implemented by following (Sukhbaatar & Fergus, 2016). We use batch size of 1024 and a learning rate of 10^{-4} for all the methods. We use one NVIDIA Tesla M40 GPU for training and another 500 CPU cores to deploy the actors to generate data.



(a) The deep neural network structure for GridNet on 5I and 3I2Z



(b) The deep neural network structure for GridNet on MAB

Figure 6. Neural network structures used for GridNet.