Reinforced Labels: Multi-Agent Deep Reinforcement Learning for Point-Feature Label Placement

1 TRAINING & HYPERPARAMETERS

We use PPO implementation from the RLLib framework [1]. Rollout workers query the current policy to determine actions and collect a new vector of observations and rewards. Collected data are assembled into training batches and shuffled. The trainer worker coordinates the rollout workers and orchestrates policy optimization. We use Adam to optimize the policy and value function parameters. Further hyperparameters are as follows:

TABLE 1 Hyperparameters

Parameter	Value
Learning rate	$1 \cdot 10^{-6}$
Batch size	2000
Mini-batch size	128
SGD epochs	10
PPO clip factor	0.25
PPO gradient clipping	-
PPO entropy coefficient	0.0
Horizon	100
γ	0.99
λ	1.0

We designed two observation vectors. First, the *mapping* vector *M* provides surroundings modalities. Second, the *self-aware* vector

S provides local information about the state of the given label

agent. Observations are mostly normalized to [-1,1] range. The intersection observation type is an exception – the value of -1

signifies the intersection of anchor, 0 defines screen bound, and

1 represents the label. We use Box2D¹ framework to cast rays

serving as LiDAR sensor to encode the agent's surroundings by 32 rays (similar to LiDAR) evenly distributed around the label

bounds that sense the range, type of the nearest intersected object (*i.e.*, label, anchor, bounds of the environment), and the mass that

TABLE 2 Observation Modalities

Mapping vector	Range	Shape
Intersection distance Intersection mass	(-1,1) (-1,1)	32 32
Number of intersected labels Intersection object type	(-1,1) (-1,0,1)	32 32
Self-aware vector	Range	Shape
Overlap area	(-1,1)	1
Overlap indicator	(-1,1)	1
Number of overlaps	(-1, 1)	1
Displacement	(-1,1)	1
Cumulative displacement	(-1,1)	1
Anchor penetration distance	(-1,1)	1
Penetration indicator	(-1, 1)	1
Number of penetrations	(-1, 1)	1
Anchor-port distance	(-1, 1)	1
Anchor-port angle	(-1, 1)	1
Anchor-port vector	(-1, 1)	2
Anchor-origin distance	(-1, 1)	1
Elapsed time steps	(-1, 1)	1

as an optional reference to a background image. Additionally, the JSON file contains a description of the labels associated with the instance, where each label is characterized by the position of its anchor and the size of its label box. Moreover, the label definition may include a text string and font configuration for each label.

```
{
    "screen": {
        "size": {"width": 2400, "height": 1600},
        "background": {"file": "background.svg"}
    },
    "labels": {
        "definition": {
            "0": {
                "text": "Brno",
                "size": [76, 20],
                "font": {
                     "size": 17,
                    "family": "Arial",
                    "style": "Regular"}
    },
    ....
}
```

3 DATASET DEFINITION

the ray went through.

OBSERVATION MODALITIES

2

The dataset comprises a collection of instances, where a JSON file represents each instance. The JSON file contains information about the screen or drawing area, including the width and height, as well

REFERENCES

 E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. I. Jordan, and I. Stoica, "Rllib: Abstractions for distributed reinforcement learning," in *Proceedings of the 35th ICML 2018*. PMLR, 2018, pp. 3059–3068.

^{1.} The framework Box2D is available at https://box2d.org.

TABLE 3

Visual comparison of examined methods applied to selected instances from the compact dataset and real-world instances of IATA airport codes with 250 anchors and CITY names with 150 anchors based on data obtained from Open Street Maps. The green dot represents the anchor (*i.e.*, point feature). The gray rectangle symbolizes the body of the label itself. The red dot describes an anchor that was not labeled by the given method. The red rectangle illustrates the dimensions of the missing label. We stress that we intentionally added all the missing labels to the visualization for illustrative purposes only, and their origins are not the outcome of the method itself.

	RFL	RAPL	PBL-A	PBL-AD
	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	<u>134 [24]9</u> 7 <u>]12]</u> 21 6 <u>9</u>	<u>134</u> 77 <u>72</u> <u>72</u> <u>72</u> <u>72</u> <u>72</u> <u>76</u> <u>76</u>	14 24 19 1 <u>7 12 12 12 16 16 16 16 16 16 16 16 16 16 16 16 16 </u>
25	8 17 15 10 18 18 18 10 18 18 18 18 10 18 18 18 18 10 18 18 18 18 18 18 18 18 18 18 18 18 18	JS J15 17 10 13 16 0 J5 J1 18	B J15 J10 17 J13 16 0 23 1 18 5 J1 1	8 15 10 17 13 16 0 23 1 16 18 5 11 1
				20 22 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2
30	1 26 9 28 16 12 23 25 3 12 12 23	26 9 28 27 16 12 27 3 14 27 14 27 17 27	200 99 727 16 3 32 25 3 3 14 921 8 8	1 20 27 27 16 1 22 25 3 13 13 12 25 3 13 12 12 12 12 12 12 12 12 12 12 12 12 12
	5 8 13 23 18 20 0 6 15 17 17 7 19 74 10	J3 J3 23 J8 20 j6 15 17 7 J9 24	js js<	15 12 13 13 13 13 13 15 10 11 15 15 15 15 15 15 15 15 15 15 15 15
	13 12 20 13 20 14 15 15 16 17 10 10 10 10 10 10 10 10 10 10 10 10 10	22	13 20 13 13 20 13 13 20 15 15 15 15 15 15 15 15 15 15 15 15 15	13 20 23 26 26 26 26 26 26 26 26 26 26 26 26 26
35	3 80 93 94 94 94 94 94 94 94 94 94 94 94 94 94	3 30 8 9 34 19 32 3 3 4 2 3 3 4 2 3 4 4 3 4 4 2 3 4 4 2 4 4 4 2 4 4 4 2 4 4 4 2 4 4 4 4 4 4 4 4 4 4 4 4 4	3 8 3 34 5 23 2 5 25 4 31 20 25 4 10 10	3 8 123 124 14 5 2 2 31 20 127 10 10 10
	22 23 24 24 24 24 24 24 24 24 24 24 24 24 24	22 28 29 24 20 24 29 24 20 20 20 20 20 20 20 20 20 20 20 20 20	37 28 24 35 72 7 2 35 22 12 23 137	27 28 24 18 18 18 18 18 18 18 18 18 18 18 18 18
	26 5 10 33 2 31 6 13 33 2 31 23 24 10 13 2 32 24 10 12 2 4 10 12 2 5 10 10 10 2 5 10 10 10 2 5 10 10 10 10 2 5 10 10 10 10 1	26 12 23 430 2 j0 53 55 5 43 25 17 7 38 32 124 j1	20 5 20 12 0 15 13 12 0 15 12 17 17 12 12 13 17	26 3 10 2 0 15 3 35 7 38 32 21 25
40	23 36 13 18 28 36 13 19 33 14 14 29 34 4 20	21 36 19 18 33 15 1 28 33 14 29 34 30 20	21 9 10 35 18 23 36 15 23 33 14 14 59 34 20 20	21 9 19 18 28 36 15 1 33 14 14 29 34 4 20
	12 27 16 39 20 44 30 30 10 </td <td>13 15 30 24 43 30 15 90 10 2 7 97</td> <td>25 5 7 12 12 12 12 12 12 12 12 12 12 12 12 12</td> <td>23 27 16 24 44 0 0 15 36 20 20 2 7 21 39 1</td>	13 15 30 24 43 30 15 90 10 2 7 97	25 5 7 12 12 12 12 12 12 12 12 12 12 12 12 12	23 27 16 24 44 0 0 15 36 20 20 2 7 21 39 1
45	13 10 34 13 10 42 14 13 12 43 14 37 26 28 13 37 35	34 18 40 14 31 32 42 43 5 25 43 43 32 5 25 16 28 41 35 6	10 13 14 13 12 34 14 14 12 12 12 12 12 13 14 15 12 10 10 10 10 10 10 15 <td>29 11 18 40 34 14 14 32 142 22 143 41 5 25 8 1.6 6 35 •</td>	29 11 18 40 34 14 14 32 142 22 143 41 5 25 8 1.6 6 35 •
	16 10 23 23 11 30 30 27 31 10 10 10 10 10 10 10 10			26 23 31 30 3 27 31 9 13 17 10 26
	42 46 3 10 34 10 12 20 48 15 24 49 20 48 7 2 57 6 6 9 13 13 9 9 10	42 78 46 10 54 12 22 20 46 15 19 40 20 20 46 24 2 37 30 6 37 36 33 33 33 5 35	12 28 46 3 10 34 32 70 10 10 15 74 19 37 30 47 36 33 6 10	42 28 46 3 10 34 12 20 48 48 15 24 19 37 30 6 47 12 13 2 6 6 36 40 32 35 35 35
50	32 3 35 5 33 7 21 25 5 30 60 22 25 14 36 0 60 76 14 36 16 1 38 45 22 24 27	32 9 33 42 33 7 22 14 25 40 20 15 16 4 11 45 36 27 16 4	32 9 5 25 33 40 7 22 16 16 16 16 16 16 16 16 16 16 16 16 16	32 9 7 22 33 40 22 14 15 4 18 38 57 40 12 12 14 14 14 14 14 14 14 14 14 14 14 14 14
CITY				
150				
ATA				
2501				

RFL RAPL 10094044 |2774 |2004 ,0000 ,000 0.20 200 300 400

 TABLE 4

 Visual comparison of RFL and RAPL methods applied to selected instances from the volume dataset.

 TABLE 5

 Visual comparison of RFL and RAPL methods applied to selected instances from the volume dataset.



PBL-A PBL-AD 5000 5000

 TABLE 6

 Visual comparison of PBL-A and PBL-AD methods applied to selected instances from the volume dataset.

 TABLE 7

 Visual comparison of PBL-A and PBL-AD methods applied to selected instances from the volume dataset.





Fig. 1. Training metrics of RFL (denoted by light blue) smoothed out using Exponential Moving Average (denoted by dark blue). Charts (a) and (b) show *mean episode reward* (same as mean return with discount factor $\gamma = 1.0$) and *loss* over within training iteration.