Perception-R1: Pioneering Perception Policy with Reinforcement Learning

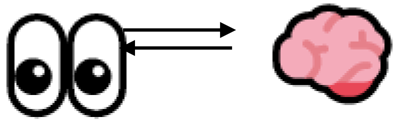
En Yu

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Highlight: PR1 serves as the first pure MLLM to surpass 30AP on COCO 2017 with only 3B size !

Part 1: Motivation

- What can RL bring to MLLM?
- Perception is a crucial prerequisite for visual reasoning.



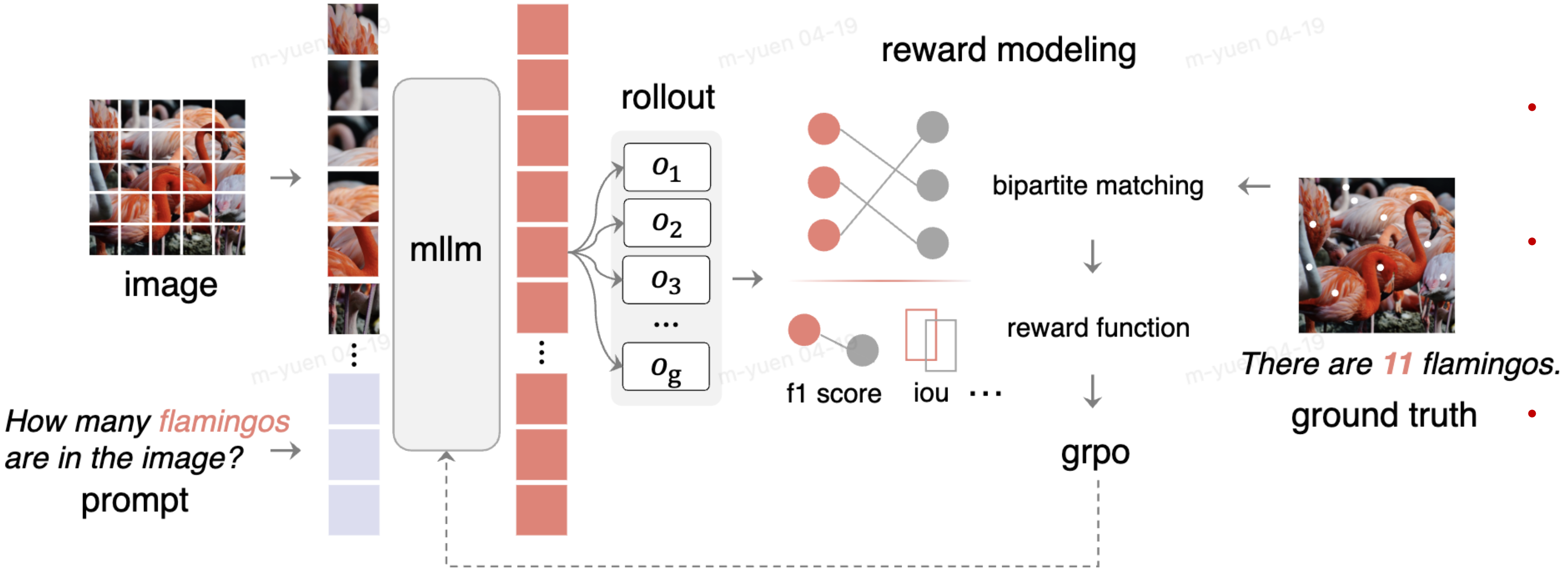
Part 2: Initial Observation

Key Properties of visual perception

- Visual perception is embodied in the objective physical world.
- Visual perception, *e.g.*, visual grounding and OCR, are mostly "single-step" direct predictions.

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 3: Perception-R1



Key Design

- Reward Modeling including format reward and answer reward where answer reward is mainly anchored in *visual discrimination, e.g., point, bbox*.
- **Reward Matching** for multi-subject reward. We use hungarian algorithm for bipartite graph matching that guarantees the optimal pairing by maximizing the overall reward.
- Reward Function. For different visual perception tasks, we tailor distinct discriminative rewards, e.g., IoU for grounding, Euclidean distance for counting.

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 3: Perception-R1

Configuration

Gradient Accmulation	Rollout G	KL Coefficient	Max Response Len	Temperature
2	8	0.04	2048	1.0

Data Recipe

tasks	datasets	Original	Used	Ratio
visual grounding	RefCOCO / RefCOCO+ / RefCOCOg	320k	5k	1.56%
OCR	PageOCR	50k	5k	10%
visual counting	PixMo-Count	1.9M	10k	0.5%
object detection	COCO2017	110k	110k	100%
overall	-	2.38M	130k	-

Highlight Work

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 4: Performance Landscape

Visual Grounding

method	size	RefCOCO $val_{@50}$	$testA_{@50}$	$testB_{@50}$	$val_{@75}$	$testA_{@75}$	$testB_{@75}$	$val_{@95}$	$testA_{@95}$	$testB_{@95}$	val_{Avg}	$testA_{Avg}$	$testB_{Avg}$
MDETR [25]	-	87.5	90.4	82.6	-	-	-	-	-	-	-	-	-
OFA [62]	-	88.4	90.6	83.3	-	-	-	-	-	-	-	-	-
LLaVA-1.5 [35]	7B	49.1	54.9	43.3	10.7	13.6	6.9	0.4	0.3	0.3	20.1	22.9	16.8
LLaVA-NeXT [36]	7B	82.5	88.4	74.0	45.7	54.8	35.6	1.9	2.6	0.7	43.4	48.6	36.8
LLaVA-OV [28]	7B	73.0	82.3	63.5	24.2	29.6	15.9	0.5	0.5	0.5	32.6	37.5	26.6
Qwen2-VL [61]	2B	86.8	89.6	82.0	77.2	80.6	70.1	33.0	35.7	26.9	65.7	68.6	59.7
Perception-R1	2B	**89.1**	**91.4**	**84.5**	**79.5**	**83.6**	**72.4**	**35.0**	**38.5**	**28.8**	**67.9**	**71.2**	**61.9**

method	size	RefCOCO+ $val_{@50}$	$testA_{@50}$	$testB_{@50}$	$val_{@75}$	$testA_{@75}$	$testB_{@75}$	$val_{@95}$	$testA_{@95}$	$testB_{@95}$	val_{Avg}	$testA_{Avg}$	$testB_{Avg}$
MDETR [25]	-	81.1	85.5	72.9	-	-	-	-	-	-	-	-	-
OFA [62]	-	81.3	87.1	74.2	-	-	-	-	-	-	-	-	-
LLaVA-1.5 [35]	7B	42.4	49.7	36.4	9.8	12.4	6.4	0.5	0.5	0.2	17.6	20.8	14.3
LLaVA-NeXT [36]	7B	74.5	84.0	64.7	41.5	51.8	30.0	1.9	2.7	1.0	39.3	46.2	31.9
LLaVA-OV [28]	7B	65.8	79.0	57.2	23.6	28.8	15.3	0.6	0.6	0.4	30.0	36.1	24.3
Qwen2-VL [61]	2B	77.1	82.5	70.1	68.7	73.8	60.0	29.4	32.3	23.0	58.4	62.9	51.0
Perception-R1	2B	**81.7**	**86.8**	**74.3**	**73.6**	**79.3**	**64.2**	**32.6**	**36.9**	**26.7**	**62.6**	**67.7**	**55.1**

method	size	RefCOCOg $val_{@50}$	$test_{@50}$	$val_{@75}$	$test_{@75}$	$val_{@95}$	$test_{@95}$	val_{Avg}	$test_{Avg}$
MDETR [25]	-	83.3	83.3	-	-	-	-	-	-
OFA [62]	-	82.2	82.3	-	-	-	-	-	-
LLaVA-1.5 [35]	7B	43.2	45.1	8.5	9.3	0.3	0.3	17.3	18.2
LLaVA-NeXT [36]	7B	77.5	77.1	40.7	39.9	1.8	1.7	40.0	39.6
LLaVA-OV [28]	7B	70.8	70.8	23.3	23.6	0.6	0.7	31.6	31.7
Qwen2-VL [61]	2B	83.3	83.1	72.7	73.0	28.9	27.9	61.6	61.3
Perception-R1	2B	**85.7**	**85.4**	**75.7**	**76.0**	**32.1**	**33.1**	**64.5**	**64.8**

OCR

	size	Edit Distance ↓ en	Edit Distance ↓ zh	F1-score ↑ en	F1-score ↑ zh	Precision ↑ en	Precision ↑ zh	Recall ↑ en	Recall ↑ zh	BLEU ↑ en	BLEU ↑ zh	METEOR ↑ en	METEOR ↑ zh
Nougat [4]	250M	25.5	-	74.5	-	72.0	-	80.9	-	66.5	-	76.1	-
DocOwl1.5 [23]	7B	25.8	-	86.2	-	83.5	-	96.2	-	78.8	-	85.8	-
GOT [65]	580M	3.5	3.8	97.2	98.0	97.1	98.2	97.3	97.8	94.7	87.8	95.8	93.9
Qwen2-VL [61]	2B	8.0	10.0	94.4	93.0	96.9	96.1	93.0	90.5	90.9	78.0	94.1	87.2
LLaVA-NeXT [36]	7B	43.0	-	64.7	-	57.3	-	88.1	-	47.8	-	58.2	-
Perception-R1	2B	**3.5**	**9.0**	**98.2**	**94.4**	**98.6**	**96.3**	**97.8**	**92.7**	**96.7**	74.6	**98.1**	**88.9**

Visual Counting & Object Detection

method	size	Visual Counting $Pixmo_{val}$	$Pixmo_{test}$
LLaVA-1.5 [35]	7B	33.3	31.0
LLaVA-1.6 [58]	7B	32.7	31.9
LLaVA-OV [28]	7B	55.8	53.7
Qwen2-VL [61]	2B	60.2	50.5
Perception-R1	2B	**78.1**	**75.6**

method	size	epoch	Object Detection AP	AP_{50}	AP_{75}
YOLOv3 [51]	-	273	27.9	49.2	28.3
Faster-RCNN [52]	-	12	35.6	55.7	37.9
DETR [6]	41M	500	42.0	62.4	44.2
Qwen2.5-VL [3]	3B	1	16.1	23.7	16.7
Perception-R1†	3B	1	**31.9**	**46.7**	**33.4**

General Image Understanding

	llm	MMBench Avg	MMVet Avg	MMStar Avg	ScienceQA Avg	SeedBench Avg	MME Cognition	MME Perception	LLaVA-Bench Avg	AI2D Avg
LLaVA1.5 [35]	Vicuna1.5-7B	62.8	32.8	32.6	65.4	60.1	302.1	1338.3	52.6	51.9
LLaVA-NeXT [36]	Vicuna1.5-7B	66.0	37.9	37.7	68.2	69.1	195. 7	1419.5	52.7	67.4
Qwen2-VL [61]	Qwen2-2B	**71.9**	45.6	**46.3**	**74.0**	72.7	418.5	1471.1	46.5	71.6
Perception-R1	Qwen2-2B	71.8	**48.9**	45.7	73.4	**73.0**	**430.0**	**1473.9**	**58.2**	**71.8**

Highlight Work

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 5: Bitter but valuable lesson

case	Visual Grounding RefCOCO	RefCOCO+	RefCOCOg	OCR PageOCR	Visual Counting $\text{Pixmo}_{\text{val}}$	$\text{Pixmo}_{\text{test}}$	Detection COCO2017
Perception-R1	89.1	81.7	85.7	98.2	78.1	75.6	31.9
w/o reward matching	-	-	-	-	77.1	75.4	23.5
w/o RL	86.8	77.1	83.3	98.2	60.2	50.5	16.1
w thinking	75.1	67.9	71.3	93.8	74.9	72.8	25.7
w/o thinking	**89.1**	**81.7**	**85.7**	98.2	**78.1**	**75.6**	28.1
RL only	**89.1**	**81.7**	**85.7**	98.2	**78.1**	**75.6**	**31.9**
SFT only	88.2	80.7	84.6	97.2	58.0	59.9	25.9
SFT+RL	88.4	80.7	85.1	**98.3**	77.1	75.4	30.8

- Reward matching enhances the explorability of multi-subject visual perception.
- Explicit thinking processes prove non-essential for contemporary visual perception.
- Perceptual perplexity dictates RL's advantage over SFT.

Highlight Work

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 5: Bitter but valuable lesson

case	Visual Grounding			OCR	Visual Counting		Detection
	RefCOCO	RefCOCO+	RefCOCOg	PageOCR	Pixmo$_{val}$	Pixmo$_{test}$	COCO2017
Perception-R1	89.1	81.7	85.7	98.2	78.1	75.6	31.9
w/o reward matching	-	-	-	-	77.1	75.4	23.5
w/o RL	86.8	77.1	83.3	98.2	60.2	50.5	16.1
w thinking	75.1	67.9	71.3	93.8	74.9	72.8	25.7
w/o thinking	**89.1**	**81.7**	**85.7**	98.2	**78.1**	**75.6**	28.1
RL only	**89.1**	**81.7**	**85.7**	98.2	**78.1**	**75.6**	**31.9**
SFT only	88.2	80.7	84.6	97.2	58.0	59.9	25.9
SFT+RL	88.4	80.7	85.1	**98.3**	77.1	75.4	30.8

- Reward matching enhances the explorability of multi-subject visual perception.
- Explicit thinking processes prove non-essential for contemporary visual perception.
- Perceptual perplexity dictates RL’s advantage over SFT.

Highlight Work

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 5: Bitter but valuable lesson

case	Visual Grounding			OCR	Visual Counting		Detection
	RefCOCO	RefCOCO+	RefCOCOg	PageOCR	$\text{Pixmo}_{\text{val}}$	$\text{Pixmo}_{\text{test}}$	COCO2017
Perception-R1	89.1	81.7	85.7	98.2	78.1	75.6	31.9
w/o reward matching	-	-	-	-	77.1	75.4	23.5
w/o RL	86.8	77.1	83.3	98.2	60.2	50.5	16.1
w thinking	75.1	67.9	71.3	93.8	74.9	72.8	25.7
w/o thinking	**89.1**	**81.7**	**85.7**	98.2	**78.1**	**75.6**	28.1
RL only	**89.1**	**81.7**	**85.7**	98.2	**78.1**	**75.6**	**31.9**
SFT only	88.2	80.7	84.6	97.2	58.0	59.9	25.9
SFT+RL	88.4	80.7	85.1	**98.3**	77.1	75.4	30.8

- Reward matching enhances the explorability of multi-subject visual perception.
- Explicit thinking processes prove non-essential for contemporary visual perception.
- Perceptual perplexity dictates RL's advantage over SFT.

Highlight Work

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 5: Bitter but valuable lesson

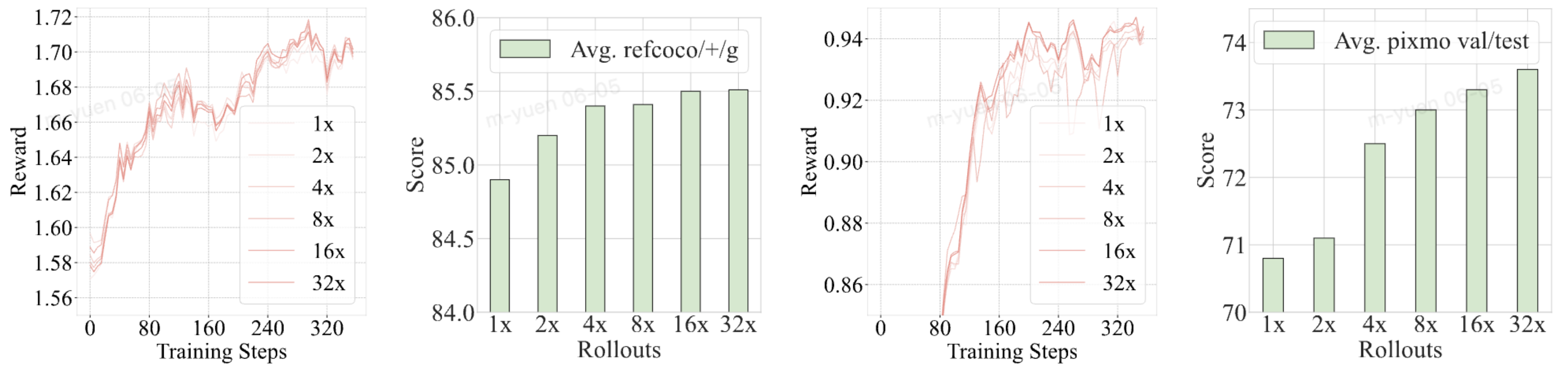
case	Visual Grounding RefCOCO	Visual Grounding RefCOCO+	Visual Grounding RefCOCOg	OCR PageOCR	Visual Counting $\text{Pixmo}_{\text{val}}$	Visual Counting $\text{Pixmo}_{\text{test}}$	Detection COCO2017
Perception-R1	89.1	81.7	85.7	98.2	78.1	75.6	31.9
w/o reward matching	-	-	-	-	77.1	75.4	23.5
w/o RL	86.8	77.1	83.3	98.2	60.2	50.5	16.1
w thinking	75.1	67.9	71.3	93.8	74.9	72.8	25.7
w/o thinking	**89.1**	**81.7**	**85.7**	98.2	**78.1**	**75.6**	28.1
RL only	**89.1**	**81.7**	**85.7**	98.2	**78.1**	**75.6**	**31.9**
SFT only	88.2	80.7	84.6	97.2	58.0	59.9	25.9
SFT+RL	88.4	80.7	85.1	**98.3**	77.1	75.4	30.8

- Reward matching enhances the explorability of multi-subject visual perception.
- Explicit thinking processes prove non-essential for contemporary visual perception.
- Perceptual perplexity dictates RL’s advantage over SFT.

Highlight Work

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 6: Property of PR1



(a) Grounding reward (b) Grounding performance (c) Counting reward (d) Counting performance

reward function	COCO2017 AP	COCO2017 AP_{50}	COCO2017 AP_{75}
format reward	-	-	-
format reward + location reward (IoU)	18.8	25.3	20.1
format reward + location reward (IoU) + cls reward	20.2	27.3	21.4
format reward + location reward (IoU) + cls reward + recall reward (F1)	27.6	42.0	28.7
format reward + location reward (IoU) + cls reward + recall reward (F1) + missing reward	**28.1**	**42.0**	**29.6**

- Scalability of PR1: Increasing rollout count enhances reward optimization and final performance.
- Reward design plays a critical role in perception policy learning.

Highlight Work

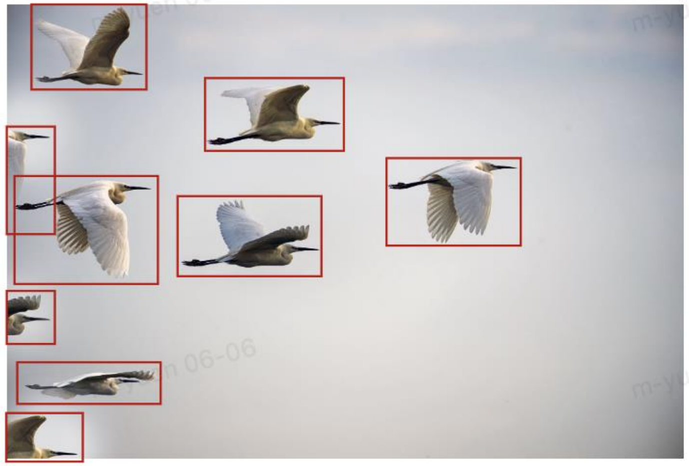
Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 7: Demo cases

Visual Counting :

Input:

Output:



Question:

<image>Output all the bounding boxes of the bird.

Perception-R1:

(259,427),(466,591)(298,172),(497,315)(558,344),(759,524)(0,273),(61,466)(30,790),(222,875)(27,9),(208,207)(0,902),(95,998)(0,637),(51,714)(15,389),(215,599)   Num boxes = 9

Highlight Work

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 7: Demo cases

Visual Grounding :

Input:

Output:

Question:

<image>Output the bounding box of the woman holding a child in the image.

Perception-R1:

The woman holding a child(157,459),(208,622)

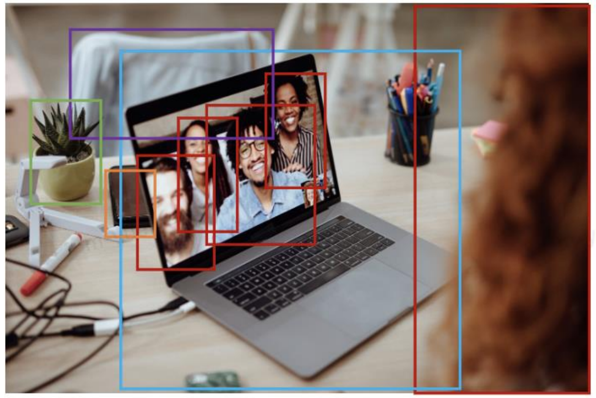
Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 7: Demo cases

Object detection :

Input:

Output:



Question:

<image>Please output bbox coordinates and names of {90 categories of COCO}.

Perception-R1:

```
[
{"bbox_2d": [435, 1, 616, 415], "label": "person"},
{"bbox_2d": [218, 112, 323, 253], "label": "person"},
{"bbox_2d": [147, 165, 225, 280], "label": "person"},
{"bbox_2d": [183, 126, 239, 242], "label": "person"},
{"bbox_2d": [276, 77, 338, 192], "label": "person"},
{"bbox_2d": [127, 54, 471, 406], "label": "laptop"},
{"bbox_2d": [108, 174, 156, 243], "label": "cell phone"},
{"bbox_2d": [26, 107, 103, 214], "label": "potted plant"},
{"bbox_2d": [64, 29, 280, 137], "label": "chair"}
]
```

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 7: Demo cases

OCR :

Input:

GEARLAB

Riding Dirty

A muddy mix of road- and mountain-bike racing, cyclocross doubles the thrill of both sports. Here's the gear to get you started.

Craft Shield Glove

Darn Tough Micro Crew

Park Tool Brush Set

Rapha Arm and Leg Warmers

Topeak SmartGauge D2

Question:

<image>OCR this image.

Perception-R1:

G E A R L A B\n\nRiding Dirty\n\nA muddy mix of road- and mountain-bike racing, cyclocross doubles the thrill of both sports. Here's the gear to get you started.\n\nby Berne Broudy\n\nCyclocross is as cool as it looks: Racers pedal drop-handlebar bikes with trail-ready tires through a course that often includes pavement, dirt, mud, and grass, as well as obstacles that force you to dismount and sling your bike over your shoulder. "All you really need is to be tough and have a good attitude and a desire to suffer a little bit," says Stu Thorne, founder and director of the Cannondale professional cyclocross team. The right gear helps, too.\n\nBest for Entry Level\n\nCannondale CAADX 105\n\nWhat you sacrifice with extra weight you make up for with a highly capable aluminum frame that'll keep you going between races. A carbon-blade fork helps eat up some of the pounding the courses can dish out. The dependable 2x11 speed Shimano 105 drivetrain powers through any grade, and TRP cable disc brakes perform well even in mud. Meanwhile, the 35cc tires and stable geometry make the CAADX a superfun commuter if you're looking for a bike that can multitask.\n\ncannondale.com $1,970\n\nBest for Experienced Cyclists\n\nSpecialized CruX Elite X1\n\nIf you already log long miles on a premium road bike, this should be your new cross rig. The carbon frame is as light as it is.\n\nA low bottom bracket and relatively slack head angle, which make it stable, helped us sail through rocks and roots and corner quickly. The tires can be run tubeless to better resist flats, and extra clearance means they spin freely when caked with mud. The CruX Elite is playful and fast — and something you won't outgrow as you collect medals.\n\nspecialized.com $3,000\n\nACCESSORIES\n\nCraft Shield Glove\n\nThe cross season typically runs from September through February, so you'll need hearty gloves like these, with a fleece lining and a waterproof base, for warmth on wet race days. craftsports.us $78\n\nDarn Tough Micro Crew\n\nUnlike other bike races, cyclocross requires you to be on foot at times. So light, strong socks are key. These aren't likely to wear out, but Darn Tough will replace them if they do. darntough.com $18\n\nPark Tool Brush Set\n\nThe mud, dirt, and grime that builds up during off-road rides can damage key components. This kit does more than just keep your bike looking fresh; it keeps it healthy, too. parktool.com $80\n\nRapha Arm and Leg Warmer\n\nThese merino layers, which have a bit of Lycra for stretch, peel off easily when the weather warms up. And they dry quickly, whether you sweat profusely or get caught in a sudden squall. rapha.cc From $70\n\nTopeak SmartGauge D2\n\nFor peak performance, adjust your tire pressure to suit the terrain. (On soft trails, lower pressure makes it grip better.) The SmartGauge makes it a snap with a fast, readable result. topeak.com $40

Perception-R1: Pioneering Perception Policy with Reinforcement Learning

Part 8: Some thoughts

- Existing visual perception tasks, *e.g.*, counting and detection, are overly simplistic, which limits the exploration space for RL. We need some *meta tasks* to unlock internal visual logic.
- How to construct effective thinking process for visual perception and reasoning, maybe think with image or agent RL?

Acknowledgements

華中科技大學
Huazhong University

StepFun

Concat Information:

email: yuen_daniel@outlook.com
homepage: https://ahnsun.github.io/