Implicit Drifting Policy: One-Step Action Generation via Conditional Expert Geometry

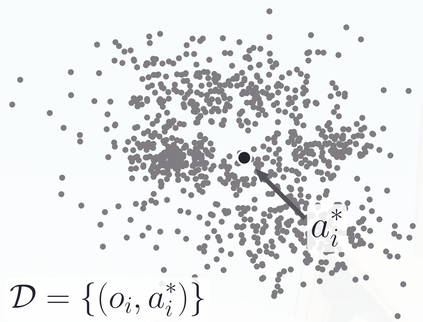
Zemin Yang[1], Yaoyu He[1], Yiming Zhong[1], Yuhao Zhang[2], Xinge Zhu[3], Yao Mu[2], Qingqiu Huang[4], Yuexin Ma[1]*

[1] ShanghaiTech University [2] Shanghai Jiao Tong University [3] The Chinese University of Hong Kong [4] Morphi Robot

Drifting learns distribution correction

Related demonstrations shape the correction during training.

Expert actions



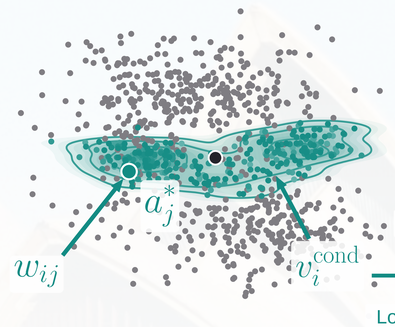
- Recorded expert actions

One expert target anchors each observation.

Related observations reveal how actions can vary around that target.

Weight by observation similarity

Conditional neighbors



w_{ij}: weight from similarity of o_i, o_j

Similar observations contribute more.

v_i^{cond} weighted action spread around the expert target, in each coordinate

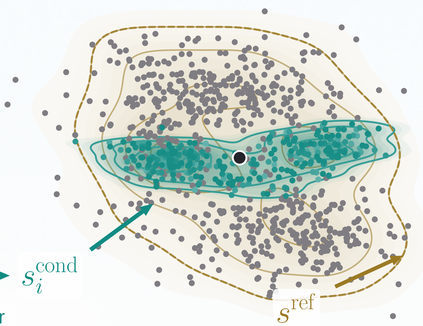
Tighter relative spread strengthens correction.

Compare inverse SD with batch

Normalized inverse SD

Local scale vector

Geometry excess



Higher normalized inverse SD: tighter relative spread.

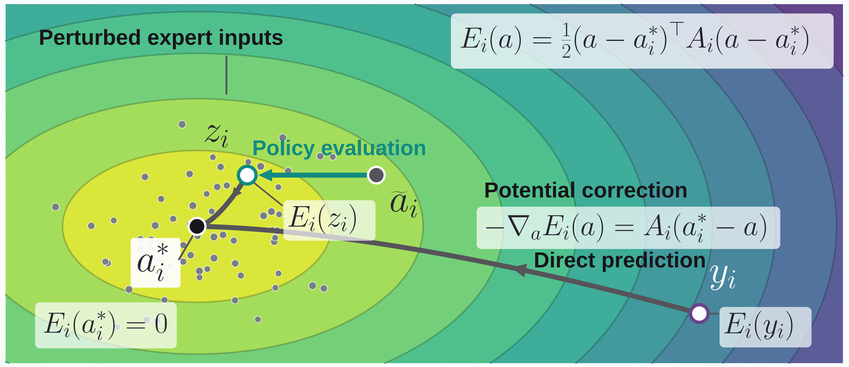
s_i^{cond} similar observations s^{ref} full batch

m_i **extra penalty for higher local consistency** than the batch. Zero keeps the base penalty.

$m_i = [s_i^{\text{cond}}/(s^{\text{ref}} + \epsilon) - 1]_+ \quad A_i = I + \mathrm{Diag}(m_i)$

Build the scalar potential

Potential supervision



Expert-proximal $z_i = f_\theta(o_i, \bar{a}_i, t_*) \quad \bar{a}_i = a_i^* + (1 - t_*)\epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, I)$

Direct $y_i = f_\theta(o_i, 0, 0)$ **Objective** $\mathcal{L} = \mathbb{E}_i[E_i(y_i) + \lambda_{\text{prox}} E_i(z_i)]$

3D manipulation across 56 tasks

81.3 ± 2.0%

Method	NFE	Adroit			DexArt				MetaWorld				Avg.
		Hammer	Door	Pen	Laptop	Faucet	Bucket	Toilet	Easy (28)	Med. (11)	Hard (5)	V. hard (5)	
DP3	10	**100±0**	58±5	53±3	83±2	40±2	25±1	74±2	91.5±0.9	74.5±5.7	54.2±1.8	71.8±4.0	79.4±2.4
Flow	10	99±1	69±3	45±3	85±0	42±4	26±3	74±3	92.2±0.7	74.6±5.0	55.2±3.0	70.8±3.0	79.9±2.2
CP	1	97±2	66±0	54±2	76±11	38±0	25±5	73±4	79.8±14	59.4±12	41.4±14	62.0±8.2	68.4±11.9
MP1	1	99±1	48±2	58±4	82±3	46±2	26±3	69±3	91.3±1.2	**75.4±5.6**	**57.6±4.6**	72.8±4.2	79.7±2.8
FlowPolicy	1	98±2	66±1	55±2	86±2	40±1	28±0	76±3	90.4±0.8	67.0±6.2	51.2±4.8	67.0±3.0	76.9±2.5
Naive Drifting	1	**100±0**	67±3	43±5	95±4	46±3	29±1	82±6	92.7±1.8	70.7±6.5	54.2±5.8	73.2±5.2	79.8±3.5
IDP	**1**	**100±0**	**67±2**	**52±4**	**96±3**	**48±1**	**30±0**	**83±6**	**93.2±0.8**	**73.6±3.9**	**56.4±4.2**	**74.2±1.4**	**81.3±2.0**

Success (%), 3 seeds. Aggregate ±: mean per-task standard deviation.

2D state and image manipulation

State 87% / Image 89%

Method	NFE	Lift		Can		Square		Transport		Tool-Hang	PushT	State avg.	Image avg.
		mh	ph	mh	ph	mh	ph	mh	ph				
DP	100	100 / 99	100 / 100	98 / 92	100 / 98	81 / 72	96 / 92	45 / 34	73 / 62	84 / 71	95 / 92	87 / 81	89 / 83
Flow	9	100 / 99	100 / 100	96 / 92	100 / 98	76 / 65	95 / 90	33 / 23	74 / 59	70 / 57	96 / 94	84 / 78	89 / 83
MIP	2	100 / 99	100 / 100	96 / 94	100 / 98	80 / 69	96 / 88	44 / 35	71 / 58	68 / 55	96 / 94	85 / 79	87 / 80
CP	1	99 / 97	100 / 100	83 / 74	99 / 96	49 / 45	88 / 82	1 / 1	32 / 24	43 / 33	94 / 91	69 / 64	79 / 70
MP1	1	100 / 99	100 / 100	96 / 87	100 / 97	77 / 62	96 / 88	24 / 14	52 / 39	76 / 63	95 / 92	82 / 74	84 / 77
Naive Drifting	1	100 / 99	100 / 100	96 / 91	99 / 96	70 / 60	97 / 90	30 / 22	68 / 58	53 / 42	95 / 91	81 / 75	87 / 78
IDP	**1**	**100 / 99**	**100 / 100**	**98 / 94**	**100 / 97**	**82 / 68**	**95 / 88**	**41 / 29**	**72 / 58**	**81 / 60**	**97 / 92**	**87 / 79**	**89 / 82**

Best / last-five success (%). mh / ph: multi-human / proficient-human.
2 architectures, 3 seeds, 50 initial conditions per seed.

Geometry and proximal training

ToolHang-State	Best (%)	Last 5 (%)
IDP	**81.0**	**60.0**
w/o proximal evaluation	40.0	25.5
w/o batch reference	78.0	54.5
Naive Drifting	53.2	42.4
Uniform neighbors	65.0	59.5
Shuffled neighbors	75.0	52.0
Absolute excess	75.0	57.0

Geometry in demonstrations

