Learning to Fly in Seconds (Parameters)

Parameter	Value
Integration Δt	$0.01\,\mathrm{s}$
Rotor 1 position	$[0.028\,\mathrm{m},\ -0.028\,\mathrm{m},\ 0\,\mathrm{m}]$
Rotor 2 position	$[-0.028\,\mathrm{m},\ -0.028\,\mathrm{m},\ 0\,\mathrm{m}]$
Rotor 3 position	$[-0.028\,\mathrm{m},\ 0.028\,\mathrm{m},\ 0\,\mathrm{m}]$
Rotor 4 position	$[0.028\,\mathrm{m},\ 0.028\,\mathrm{m},\ 0\,\mathrm{m}]$
Rotor 1 thrust direction	[0, 0, 1]
Rotor 2 thrust direction	[0, 0, 1]
Rotor 3 thrust direction	[0, 0, 1]
Rotor 4 thrust direction	[0, 0, 1]
Rotor 1 torque direction	[0, 0, −1]
Rotor 2 torque direction	[0, 0, 1]
Rotor 3 torque direction	[0, 0, −1]
Rotor 4 torque direction	[0, 0, 1]
K_f	3.16×10^{-10}
$\frac{K_d}{K_f}$	0.005 964 552
Vehicle mass	$0.027\,\mathrm{kg}$
Gravity	$[0,\ 0,\ -9.81\,\mathrm{m\,s^{-2}}]$
I_{xx}	$3.85 \times 10^{-6}\,\mathrm{kg\,m^2}$
I_{yy}	$3.85 \times 10^{-6}\,\mathrm{kg\,m^2}$
I_{zz}	$5.9675 \times 10^{-6}\,\mathrm{kg\,m^2}$
τ (RPM time constant)	$0.15\,\mathrm{s}$
RPM range	[0, 21 702]

Table 1: Parameters: Quadrotor dynamics (Crazyflie)

Parameter	Value	Description
	$C_{\text{init},*}$	
C_{rs}	2	Reward bonus for survival
C_{rp}	2.5	Position weight
C_{rq}	2.5	Orientation weight
C_{rv}	0.005	Linear velocity weight
$C_{r\omega}$	0	Angular velocity weight
C_{ra}	0.005	Action weight
C_{rab}	0.334	Action baseline
	$C_{\text{target},*}$	
C_{rs}	2	Reward bonus for survival
C_{rp}	20	Position weight
C_{rq}	2.5	Orientation weight
C_{rv}	0.5	Linear velocity weight
$C_{r\omega}$	0	Angular velocity weight
C_{ra}	0.5	Action weight
C_{rab}	0.334	Action baseline
	Curriculum Parameters	
N_C	100 000	interval of the application of multiplicative steps (curriculum)
C_{cp}	1.2	curriculum: position factor
C_{cpl}	20	curriculum: position weight limit
C_{cv}	1.4	curriculum: linear velocity factor
C_{cvl}	0.5	curriculum: linear velocity weight limit
C_{ca}	1.4	curriculum: action factor
C_{cal}	0.5	curriculum: action weight limit

Table 2: Parameters: Reward function and curriculum

Parameter	Value	Description
Guidance	0.1	probability of spawning at the origin position and at zero angle but with random linear and angular velocity
Position	Uniform($-0.2\,\text{m}$, $0.2\,\text{m}$)	
Orientation	Uniform(SO3) s.t. $\alpha \leq 90°$	
Linear Velocity	Uniform($-1\,\text{m}\,\text{s}^{-1}$, $1\,\text{m}\,\text{s}^{-1}$)	
Angular Velocity	Uniform($-1\,\text{rad}\,\text{s}^{-1}$, $1\,\text{rad}\,\text{s}^{-1}$)	
RPM	Uniform($\frac{21702}{2}$, $\frac{21702}{2}$)	
Force disturbance	Uniform($\frac{-0.027 \cdot 9.81}{20}$, $\frac{0.027 \cdot 9.81}{20}$)	$\frac{1}{20}$ of the hovering thrust
Torque disturbance	Uniform($\frac{-0.027 \cdot 9.81}{10000}$, $\frac{0.027 \cdot 9.81}{10000}$)	$\frac{1}{10000}$ of the hovering thrust

Table 3: Parameters: Initial state distribution

Parameter	Value
Observation noise position (std)	0.001
Observation noise orientation (std)	0.001
Observation noise linear velocity (std)	0.002
Observation noise angular velocity (std)	0.002

Table 4: Parameters: Observation noise

Parameter	Value
Max position error	0.6 m
Max linear velocity error	$1000\,\mathrm{m\,s^{-1}}$
Max angular velocity error	$1000\,\mathrm{rad\,s^{-1}}$

Table 5: Parameters: Termination conditions

Parameter	Value
Actor layers	[64, 64]
Actor activation function	Tanh
Actor output activation function	Tanh
Actor batch size	256
Actor warmup steps (before training)	30 000
Actor training interval	20
Actor Polyak factor	0.995
N_H (action history length in the actor observation)	32
Critic layers	[64, 64]
Critic activation function	Tanh
Critic output activation function	Identity
Critic batch size	256
Critic training interval	20
Critic warmup steps (before training)	15 000
Actor Polyak factor	0.995
Target action noise clip	0.5
Target action noise	0.5
γ	0.99
Replay buffer capacity	300 000, 3 000 000
Environment step limit	500
Exploration noise std	0.5
Exploration noise decay start (step)	500 000
Exploration noise decay interval (steps)	100 000
Exploration noise decay factor	0.9

Table 6: Parameters: Asymmetric Actor-Critic RL setup