A. Experiment Details

The corresponding hyperparameters used in Atari experiments are shown in Table 4 and Table 5. We follow Kostrikov et al. (2020) to use data augmentation techniques that consist of a simple random shift which has been shown effective in visual domain RL. Specifically, the images are padded each side by 4 pixels (by repeating boundary pixels) and then select a random 84×84 crop, yielding the original image. This procedure is repeated every time an image is sampled from the replay buffer.

We also use the same generalized policy improvement(GPI) (Barreto et al., 2017; 2020) as in VISR with the number of polices 10. GPI is also used in VISR to ensure a fair comparison. Per common practice, we average performance of our agent over 5 random seeds. The evaluation is done for 125K environment steps at the end of training for 100K environment steps.

The ablated variant APS w/o shared encoder follows APT (Liu & Abbeel, 2021) but the output dimension of the neural encoder f is decreased to 5 in order to match the default APS. The projection network in contrastive learning is a two-layer MLP with hidden size of 128 and output size of 64. We also use the same temperature and other hyperparamters as APT for the ablation study.

B. Scores Breakdown on 57 Atari games

A comparison between APS and baselines on each individual game of the 57 Atari game suite is shown in Table 6. APS achieves super-human performance on 15 games, compared to a maximum of 12 for any previous methods and achieves scores significantly higher than any previous methods.

Table 4: Hyper-parameters for RL.

Parameter	Setting
Terminal on loss of life	True
Reward clipping (fine-tuning phase)	$[-1, 1]$
Data augmentation	Random shifts and Intensity
Grey-scaling	True
Observation down-sampling	84×84
Frames stacked	4
Action repetitions	4
Max frames per episode	108k
Update	Double Q
Target network: update period	100
Discount factor	0.99
Minibatch size	32
ψ, ϕ optimizer	Adam
ψ, ϕ optimizer (pre-training phase): learning rate	0.0001
ψ, ϕ optimizer (fine-tuning phase): learning rate	0.001
ψ, ϕ optimizer: β_1	0.9
ψ, ϕ optimizer: β_2	0.999
ψ, ϕ optimizer: ϵ	0.00015
Max gradient norm	10
Training steps (fine-tuning phase)	60K
Task identity steps (fine-tuning phase)	40K
Training steps (pre-training phase)	5M
Evaluation steps	125K
Min replay size for sampling	1600
Memory size	Unbounded
Replay period every	1 step
Multi-step return length	10
ψ network: channels	$32, 64, 64$
ψ network: filter size	$8 \times 8, 4 \times 4, 3 \times 3$
ψ network: stride	$4, 2, 1$
ψ network: hidden units	512
ψ Non-linearity	`ReLU`
Exploration	ϵ-greedy
ϵ-decay	2500

Table 5: Hyper-parameters for Learning ϕ.

Parameter	Setting
Value of k	search in $\{3, 5, 10\}$
ϕ network: channels	$32, 64, 64$
ϕ network: filter size	$8 \times 8, 4 \times 4, 3 \times 3$
ϕ network: stride	$4, 2, 1$
ϕ network: hidden units	512
ϕ network Non-linearity	`ELU`
FC hidden size	1024
Output size	5

Table 6: Comparison of raw scores of each method on Atari games. Results are averaged over five random seeds. @N represents the amount of RL interaction utilized at fine-tuning phase.

Game	Random	Human	VISR	APT	APS (ours)
Alien	227.8	7127.7	364.4	**2614.8**	934.9
Amidar	5.8	1719.5	186.0	**211.5**	188.4
Assault	222.4	742.0	**1209.1**	891.5	413.3
Asterix	210.0	8503.3	**6216.7**	185.5	1159.5
Asteroids	7191	47388.7	**4443.3**	678.7	1519.7
Atlantis	12850.0	29028.1	140542.8	40231.0	**18920.0**
Bank Heist	14.2	753.1	71.3	**416.7**	262.7
Battle Zone	2360.0	37187.5	7072.7	7065.1	**26920.1**
Beam Rider	363.9	16826.5	1741.9	3487.2	**4981.2**
Berzerk	123.7	2630.4	490.0	**493.4**	387.4
Bowling	23.1	160.7	21.2	-56.5	**56.5**
Boxing	0.1	12.1	13.4	21.3	**36.3**
Breakout	1.7	30.5	17.9	10.9	**19.1**
Centipede	2090.9	12017.1	**7184.9**	6233.9	3915.7
Chopper Command	811.0	7387.8	800.8	317.0	**2517.0**
Crazy Climber	10780.5	23829.4	49373.9	44128.0	**67328.1**
Defender	2874.5	18688.9	15876.1	5927.9	**19921.5**
Demon Attack	107805	35829.4	**8994.9**	6871.8	7989.0
Double Dunk	-18.6	-16.4	-22.6	-17.2	**-8.0**
Enduro	0.0	860.5	-3.1	-0.3	**216.8**
Fishing Derby	-91.7	-38.7	-93.9	-5.6	**-2.1**
Freeway	0.0	29.6	-12.1	**29.9**	27.1
Frostbite	65.2	4334.7	230.9	**1796.1**	496.1
Gopher	257.6	2412.5	498.6	**2190.4**	2590.4
Gravitar	173.0	3351.4	328.1	**542.0**	487.0
Hero	1027.0	30826.4	663.5	6789.1	**12189.3**
Ice Hockey	-11.2	0.9	-18.1	-30.1	**-11.3**
Jamesbond	29.0	302.8	484.4	356.1	**622.3**
Kangaroo	52.0	3035.0	1761.9	412.0	**5280.1**
Krull	1598.0	2665.5	3142.5	2312.0	**4496.0**
Kung Fu Master	258.5	22736.3	16754.9	17357.0	13112.1
Montezuma Revenge	0.0	4753.3	0.0	147.0	**211.0**
Ms Pacman	307.3	6951.6	558.5	**2527.1**	2092.3
Name This Game	2292.3	8049.0	2605.8	1387.2	**6898.8**
Phoenix	761.4	7242.6	**7162.2**	3874.2	6871.8
Pitfall	-229.4	6463.7	-370.8	-12.8	**-6.2**
Pong	-20.7	14.6	-26.2	-8.0	**12.5**
Private Eye	24.9	69571.3	98.3	96.1	**117.9**
Qbert	163.9	13455.0	666.3	17671.2	**19271.4**
Riverraid	1338.5	17118.0	5422.2	4671.0	**10521.3**
Road Runner	11.5	7845.0	**6146.7**	4782.1	5919.0
Robotank	2.2	11.9	10.0	**13.7**	12.6
Seaquest	68.4	42054.7	706.6	2116.7	**4209.7**
Skiing	-17098.1	-4336.9	**-19692.5**	-38434.1	-9102.1
Solaris	1236.3	12326.7	**1921.5**	841.8	1095.4
Space Invaders	148.0	1668.7	**9741.0**	3687.2	3693.8
Star Gunner	664.0	10250.0	25827.5	8717.0	**42970.0**
Surround	-10.0	6.5	-15.5	**-2.5**	-5.8
Tennis	-23.8	-8.3	0.7	1.2	**8.7**
Time Pilot	3568.0	5229.2	4503.6	2567.0	**4586.5**
Tutankham	11.4	167.6	50.7	**124.6**	45.6
Up N Down	533.4	11693.2	**10037.6**	8289.4	4911.9
Venture	0.0	1187.5	-1.7	**231.0**	136.0
Video Pinball	0.0	17667.9	**35120.3**	2817.1	154414.1
Wizard Of Wor	563.5	4756.5	853.3	1265.0	**1732.1**
Yars Revenge	3092.9	54576.9	5543.5	1871.5	**6539.5**
Zaxxon	32.5	9173.3	897.5	3231.0	**5819.2**
Mean Human-Norm'd	0.000	1.000	68.42	47.78	**103.04**
Median Human-Norm'd	0.000	1.000	9.41	33.41	**39.23**
#Superhuman	0	N/A	11	12	**15**