

Test-Time Visual In-Context Tuning

—Supplementary Material

Table 1. **Effect on in-domain test samples.** We perform VICT on clean test images for each task. VICT does not hurt or only slightly hurts the original performance.

	depth estimation NYUv2			semantic seg. ADE-20K	panoptic seg. COCO	denoising SIDDD		deraining 5 datasets		enhance. LoL	
	RMSE ↓	A.Rel ↓	d1 ↑	mIoU ↑	PQ ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Painter	0.288	0.080	0.950	49.9	43.4	38.88	0.954	29.49	0.868	22.40	0.872
VICT	0.292	0.080	0.949	49.9	43.6	38.38	0.954	29.34	0.867	22.25	0.872

Table 2. **Results on more VICL baselines and unseen tasks.** VICT can generalize well to more VICL baselines and out-of-domain vision tasks at test time.

	surface normal estimation Taskonomy					foreground seg. DUTS	deblurring GoPro		colorization ImageNet	
	mean ↓	median ↓	11.25°↑	22.5°↑	30°↑	mIoU ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
MAE-VQGAN	28.8	26.8	18.6	41.4	54.5	38.1	14.05	0.559	13.60	0.481
w/ VICT	25.1	23.1	21.0	48.0	66.3	41.1	14.27	0.569	14.17	0.509
Painter	37.2	36.1	2.6	20.0	44.2	48.5	23.88	0.818	18.14	0.809
w/ VICT	23.1	21.0	26.4	52.9	69.9	54.1	24.50	0.829	20.04	0.873

A. Effect on In-Domain Test Samples

In the main text, we apply VICT when a distribution shift occurs at test time. It is interesting to ask whether VICT works when the training and test distributions are the same. To study this, we perform VICT on clean test images for each task. As shown in Table 1, we observe no further performance gains or slight performance degradations compared with the Painter baseline. Since the Painter model has been well pre-trained on clean images, further applying VICT cannot extract additional information from data belonging to the training distribution and might instead lead to slight overfitting in some cases. In practice, one may rely on prior domain knowledge about the deployed environment to determine whether to apply VICT. Even if such knowledge is unavailable, in most cases, VICT does not hurt or only slightly hurts the original performance, but can significantly improve on many new distributions.

B. Results on More Baselines and Tasks

In the main text, we mainly consider Painter as our baseline. To further demonstrate the generalizability of our method,

we apply VICT to another VICL baseline, *i.e.*, MAE-VQGAN [1]. Besides, we also apply VICT to more unseen tasks, including surface normal estimation on Taskonomy [5], foreground segmentation on DUTS [4], image deblurring on GoPro [3], and colorization on ImageNet [2]. We randomly sample 1000 images from their respective test or validation sets for evaluation. We sweep the learning rate for different tasks and train each test sample for 100 steps. As shown in Table 2, VICT consistently outperforms each baseline on different tasks by clear margins.

C. Results on Other Severity Levels

Table 3-6 present the results on the other four corruption levels. In all experiments, we perform VICT using the same hyper-parameters as described in Section 4.1 in the main text without specialized tuning, except that we train each test sample for 20 steps (instead of 60, for faster experiments). Note that using the same hyper-parameters across different corruptions and levels can be sub-optimal. Nevertheless, VICT outperforms Painter regardless of settings, tasks, and corruption levels.

Table 3. **System-level comparison on six representative vision tasks across 15 corruptions.** Results are on corruption level 1.

method	brigh	cont	defoc	elast	fog	frost	gauss	glass	impul	jpeg	motn	pixel	shot	snow	zoom	avg
(a) depth estimation NYUv2-C (A.Rel ↓)																
<i>zero-shot setting:</i>																
Painter	0.094	0.089	0.349	0.096	0.093	0.120	0.096	0.250	0.109	0.169	0.103	0.100	0.093	0.209	0.154	0.142
VICT	0.083	0.101	0.295	0.085	0.105	0.120	0.089	0.202	0.095	0.133	0.089	0.085	0.085	0.209	0.127	0.127
<i>one-shot setting:</i>																
Painter	0.093	0.089	0.351	0.097	0.093	0.136	0.099	0.248	0.113	0.170	0.101	0.101	0.097	0.270	0.152	0.147
VICT	0.084	0.093	0.098	0.084	0.096	0.114	0.090	0.088	0.094	0.089	0.087	0.084	0.086	0.130	0.117	0.096
(b) semantic segmentation ADE20K-C (mIoU ↑)																
<i>zero-shot setting:</i>																
Painter	48.9	46.5	45.3	45.8	45.9	43.1	46.4	45.4	46.1	48.4	47.8	48.9	46.2	43.8	31.4	45.3
VICT	49.1	46.6	45.6	45.9	46.4	43.4	46.9	45.5	46.7	48.3	47.6	48.9	47.5	43.8	31.9	45.6
<i>one-shot setting:</i>																
Painter	48.9	46.4	45.1	45.9	45.9	43.2	46.6	45.6	45.9	48.4	47.7	48.9	46.3	42.8	31.5	45.3
VICT	49.2	46.7	45.3	46.1	46.0	44.4	47.6	46.0	47.3	48.5	48.0	48.7	47.5	43.3	31.9	45.8
(c) panoptic segmentation COCO-C (PQ ↑)																
<i>zero-shot setting:</i>																
Painter	42.2	40.7	39.5	40.3	40.0	38.1	40.2	39.5	39.2	40.2	41.0	41.2	40.4	37.6	27.7	39.2
VICT	42.9	41.2	40.0	40.7	40.5	38.6	40.8	40.3	39.8	40.8	41.3	42.0	40.9	38.2	28.2	39.7
<i>one-shot setting:</i>																
Painter	42.1	40.3	39.3	40.1	39.5	38.0	40.0	39.4	38.8	39.4	40.9	41.0	40.2	37.2	27.5	38.9
VICT	42.8	40.9	39.7	40.6	40.3	38.6	40.6	40.0	39.4	40.5	41.2	41.8	40.9	37.6	27.8	39.5
(d) denoising SIDD-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	22.53	28.35	34.64	36.06	18.72	12.34	35.63	33.61	34.35	32.88	33.88	37.36	37.01	16.19	31.37	29.66
VICT	22.90	28.50	34.49	35.87	18.85	12.50	35.59	33.48	34.42	32.86	33.75	37.24	37.06	16.27	31.36	29.68
<i>one-shot setting:</i>																
Painter	23.20	28.36	34.95	36.09	18.56	12.98	35.87	33.89	34.61	32.60	33.98	37.36	37.09	16.88	31.38	29.85
VICT	28.07	27.82	35.04	36.27	17.68	15.77	35.65	34.46	35.21	33.09	34.43	37.00	36.45	21.14	31.92	30.67
(e) deraining Rain-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	22.94	18.04	24.17	24.39	15.95	14.84	27.50	23.91	26.94	28.16	23.61	27.62	27.22	17.95	19.62	22.86
VICT	23.03	18.10	24.23	24.39	16.00	14.96	27.63	23.93	27.02	28.12	23.62	27.58	27.37	18.11	19.67	22.92
<i>one-shot setting:</i>																
Painter	23.34	18.20	24.10	24.39	15.97	15.24	27.43	23.94	26.87	28.16	23.65	27.63	27.21	18.30	19.65	22.94
VICT	27.94	20.48	24.53	24.49	16.18	21.35	27.63	24.13	27.15	28.14	23.92	27.59	27.51	23.86	20.02	24.33
(f) low-light enhancement LoL-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	18.91	17.58	21.50	21.25	12.43	14.20	17.00	20.79	18.57	19.07	20.78	21.92	21.51	12.45	18.42	18.43
VICT	19.00	17.75	21.53	21.23	12.53	14.26	17.48	20.76	19.25	19.11	20.83	21.94	21.47	12.58	18.43	18.54
<i>one-shot setting:</i>																
Painter	19.27	17.84	21.50	21.24	12.33	14.20	16.69	20.81	16.60	19.08	20.78	21.92	21.55	11.62	18.41	18.26
VICT	20.72	19.65	21.26	21.16	12.53	14.93	19.74	20.79	19.41	19.30	20.74	21.90	21.56	15.45	18.37	19.17

Table 4. **System-level comparison on six representative vision tasks across 15 corruptions.** Results are on corruption level 2.

method	brigh	cont	defoc	elast	fog	frost	gauss	glass	impul	jpeg	motn	pixel	shot	snow	zoom	avg
(a) depth estimation NYUv2-C (A.Rel ↓)																
<i>zero-shot setting:</i>																
Painter	0.101	0.097	0.447	0.097	0.100	0.255	0.104	0.333	0.107	0.214	0.114	0.093	0.097	0.939	0.168	0.218
VICT	0.088	0.114	0.390	0.087	0.113	0.255	0.096	0.281	0.099	0.173	0.096	0.084	0.091	0.939	0.143	0.203
<i>one-shot setting:</i>																
Painter	0.100	0.097	0.457	0.098	0.100	0.376	0.117	0.330	0.131	0.212	0.112	0.094	0.104	1.412	0.168	0.261
VICT	0.086	0.099	0.106	0.088	0.102	0.155	0.099	0.095	0.106	0.096	0.092	0.083	0.095	0.625	0.136	0.138
(b) semantic segmentation ADE20K-C (mIoU ↑)																
<i>zero-shot setting:</i>																
Painter	47.1	45.0	42.6	43.3	44.2	35.7	44.1	43.3	43.5	47.5	44.9	48.3	43.7	29.2	25.8	41.9
VICT	47.5	45.5	42.9	42.8	45.2	35.9	44.2	43.4	43.6	47.4	45.6	48.6	43.9	30.8	26.1	42.2
<i>one-shot setting:</i>																
Painter	47.1	45.0	42.4	43.3	44.2	35.3	43.8	43.4	43.0	47.5	44.7	48.2	43.9	23.2	26.0	41.4
VICT	47.8	45.4	43.4	43.1	44.6	36.6	44.7	43.6	43.8	47.6	44.9	48.6	44.3	32.7	26.7	42.5
(c) panoptic segmentation COCO-C (PQ ↑)																
<i>zero-shot setting:</i>																
Painter	41.5	39.5	37.6	38.5	38.8	33.0	38.5	37.4	37.3	38.9	38.8	41.2	38.4	30.4	23.4	36.9
VICT	42.2	40.0	38.0	39.1	39.5	33.6	39.1	37.9	37.8	39.6	39.0	41.8	39.0	31.4	23.5	37.4
<i>one-shot setting:</i>																
Painter	41.4	38.9	37.2	38.4	38.2	32.5	38.0	37.2	36.7	38.2	38.7	41.0	38.0	29.2	23.0	36.4
VICT	42.0	39.6	37.6	39.0	39.1	33.3	38.6	37.8	37.3	39.1	38.8	41.7	38.6	31.5	23.5	37.2
(d) denoising SIDD-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	16.32	27.50	34.50	35.34	17.96	9.84	31.31	33.58	29.81	30.56	32.61	37.48	34.69	11.97	30.61	27.61
VICT	16.54	27.64	34.40	35.09	18.13	9.90	31.34	33.40	29.74	31.93	32.47	37.36	34.73	11.93	30.57	27.68
<i>one-shot setting:</i>																
Painter	17.05	27.51	34.70	35.33	17.63	10.39	32.87	33.94	30.78	30.23	32.68	37.46	35.87	12.62	30.61	27.98
VICT	21.52	27.20	34.86	35.70	16.42	11.59	33.24	34.34	32.05	32.67	33.48	37.14	35.54	15.90	31.10	28.85
(e) deraining Rain-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	17.95	16.75	22.96	23.22	15.10	13.23	26.36	23.18	25.57	27.68	21.60	27.14	25.62	14.45	18.75	21.30
VICT	18.10	16.81	23.03	23.27	15.15	13.34	26.54	23.28	25.77	27.68	21.67	27.14	25.91	14.63	18.80	21.41
<i>one-shot setting:</i>																
Painter	18.39	16.85	22.88	23.22	15.05	13.66	26.25	23.20	25.43	27.69	21.60	27.16	25.54	14.80	18.79	21.37
VICT	24.52	18.27	23.51	23.55	15.19	19.15	26.68	23.89	26.07	27.67	21.99	27.18	26.16	21.09	19.15	22.94
(f) low-light enhancement LoL-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	17.51	16.58	21.05	20.83	12.35	13.26	14.23	20.80	15.96	18.91	19.91	21.78	20.37	12.77	17.72	17.60
VICT	17.66	16.72	21.16	20.77	12.64	13.24	14.68	20.89	16.58	18.88	19.95	21.82	20.39	12.88	17.75	17.73
<i>one-shot setting:</i>																
Painter	17.89	16.86	21.07	20.83	12.04	13.30	14.07	20.83	14.57	18.92	19.93	21.79	20.39	12.09	17.71	17.49
VICT	19.29	19.16	20.91	20.81	12.55	13.20	16.16	20.75	17.03	18.59	19.94	21.77	20.74	13.89	17.74	18.17

Table 5. **System-level comparison on six representative vision tasks across 15 corruptions.** Results are on corruption level 3.

method	brigh	cont	defoc	elast	fog	frost	gauss	glass	impul	jpeg	motn	pixel	shot	snow	zoom	avg
(a) depth estimation NYUv2-C (A.Rel ↓)																
<i>zero-shot setting:</i>																
Painter	0.109	0.112	0.550	0.099	0.106	0.423	0.114	0.390	0.112	0.248	0.127	0.134	0.102	1.257	0.164	0.270
VICT	0.093	0.133	0.491	0.092	0.122	0.423	0.111	0.332	0.109	0.204	0.105	0.105	0.100	1.257	0.155	0.255
<i>one-shot setting:</i>																
Painter	0.107	0.116	0.569	0.100	0.107	0.608	0.212	0.391	0.222	0.247	0.125	0.137	0.128	1.528	0.164	0.317
VICT	0.088	0.117	0.125	0.092	0.108	0.185	0.129	0.125	0.134	0.105	0.098	0.091	0.108	0.795	0.162	0.164
(b) semantic segmentation ADE20K-C (mIoU ↑)																
<i>zero-shot setting:</i>																
Painter	46.4	41.1	36.7	39.8	42.0	27.9	39.8	33.2	40.1	45.8	41.2	46.3	39.5	24.1	21.2	37.7
VICT	46.4	41.4	37.2	40.2	42.3	28.7	39.9	34.1	41.1	45.7	41.0	46.7	39.9	28.4	21.9	38.3
<i>one-shot setting:</i>																
Painter	46.5	41.4	36.5	39.9	42.4	26.6	38.5	33.4	39.1	45.7	41.2	46.4	38.9	18.9	21.3	37.1
VICT	46.5	42.1	37.4	40.7	42.8	28.8	39.4	34.6	40.7	46.2	41.1	47.0	40.2	33.9	22.4	38.9
(c) panoptic segmentation COCO-C (PQ ↑)																
<i>zero-shot setting:</i>																
Painter	40.7	36.9	33.0	35.9	37.5	29.3	35.7	29.7	35.6	38.1	35.2	39.5	35.8	28.3	19.0	34.0
VICT	41.3	37.4	33.5	36.3	38.0	29.8	36.2	30.3	36.0	38.7	35.4	40.0	36.4	29.8	19.3	34.6
<i>one-shot setting:</i>																
Painter	40.7	36.4	32.7	35.7	37.1	28.7	34.8	29.6	34.7	37.4	35.2	39.3	35.0	26.5	18.5	33.5
VICT	41.1	36.9	33.2	36.3	37.6	29.7	35.4	30.4	35.4	38.2	35.2	39.7	35.8	30.8	19.0	34.3
(d) denoising SIDD-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	12.83	26.95	34.34	34.39	17.35	8.90	25.99	32.31	26.27	29.35	31.48	36.30	31.45	11.63	30.38	25.99
VICT	13.00	26.95	34.18	34.13	17.48	8.94	26.10	32.26	26.24	30.96	31.32	36.18	31.48	11.59	30.33	26.08
<i>one-shot setting:</i>																
Painter	13.47	26.97	34.37	34.42	17.06	9.36	27.79	32.57	27.93	29.07	31.48	36.28	32.36	12.14	30.34	26.37
VICT	16.80	26.86	34.10	35.18	15.86	11.07	29.68	33.13	29.28	31.83	32.10	36.10	33.31	15.85	30.67	27.45
(e) deraining Rain-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	14.69	15.62	21.22	21.91	14.33	12.27	24.59	20.35	24.50	27.31	20.04	25.35	23.77	14.93	18.57	19.96
VICT	14.81	15.69	21.28	22.02	14.39	12.40	24.90	20.47	24.80	27.29	20.11	25.39	24.20	15.17	18.61	20.10
<i>one-shot setting:</i>																
Painter	15.09	15.64	21.20	21.92	14.28	12.66	24.47	20.42	24.27	27.32	20.11	25.40	23.53	15.24	18.60	20.01
VICT	20.81	16.43	21.92	22.55	14.57	18.01	25.33	21.43	25.26	27.37	20.71	25.58	24.38	20.93	19.01	21.62
(f) low-light enhancement LoL-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	17.00	15.78	20.14	20.43	10.93	12.45	12.64	19.55	14.44	18.38	19.02	21.36	20.15	12.36	17.40	16.80
VICT	17.12	15.85	20.17	20.42	10.93	12.44	12.58	19.60	14.60	18.31	19.07	21.36	20.43	12.56	17.49	16.86
<i>one-shot setting:</i>																
Painter	17.36	16.00	20.16	20.43	10.56	12.88	12.59	19.58	13.99	18.28	19.07	21.36	19.97	11.66	17.38	16.75
VICT	18.34	17.85	20.19	20.49	11.44	13.09	13.07	19.59	16.64	17.95	19.11	21.38	20.73	13.35	17.47	17.38

Table 6. **System-level comparison on six representative vision tasks across 15 corruptions.** Results are on corruption level 4.

method	brigh	cont	defoc	elast	fog	frost	gauss	glass	impul	jpeg	motn	pixel	shot	snow	zoom	avg
(a) depth estimation NYUv2-C (A.Rel ↓)																
<i>zero-shot setting:</i>																
Painter	0.118	0.148	0.646	0.102	0.111	0.449	0.141	0.440	0.139	0.325	0.154	0.178	0.120	1.500	0.185	0.317
VICT	0.100	0.170	0.582	0.098	0.128	0.449	0.137	0.380	0.139	0.275	0.128	0.136	0.120	1.500	0.178	0.301
<i>one-shot setting:</i>																
Painter	0.115	0.185	0.668	0.103	0.112	0.696	0.470	0.443	0.568	0.324	0.157	0.180	0.263	1.699	0.186	0.411
VICT	0.092	0.164	0.181	0.098	0.114	0.216	0.173	0.174	0.247	0.127	0.114	0.103	0.147	1.017	0.189	0.210
(b) semantic segmentation ADE20K-C (mIoU ↑)																
<i>zero-shot setting:</i>																
Painter	43.4	28.4	32.1	35.6	40.9	27.8	33.7	30.5	34.0	43.1	35.2	43.5	31.9	16.3	17.5	32.9
VICT	44.0	28.9	32.2	36.0	41.0	28.6	34.0	30.9	34.7	43.0	36.0	43.6	33.0	20.6	17.8	33.6
<i>one-shot setting:</i>																
Painter	44.1	28.6	31.6	35.8	41.2	27.1	30.9	30.8	30.7	43.0	35.4	43.5	30.5	12.3	17.6	32.2
VICT	44.4	31.5	32.7	35.6	41.2	29.4	33.0	31.6	33.3	43.6	36.3	44.3	32.8	19.3	18.6	33.8
(c) panoptic segmentation COCO-C (PQ ↑)																
<i>zero-shot setting:</i>																
Painter	39.5	28.5	29.4	33.2	36.7	28.7	31.3	27.7	31.2	35.8	30.6	37.0	30.7	23.7	15.6	30.6
VICT	40.2	29.2	29.6	33.6	37.2	29.2	31.8	28.3	31.5	36.3	30.7	37.5	31.0	25.1	15.9	31.1
<i>one-shot setting:</i>																
Painter	39.6	28.0	28.8	33.1	36.2	27.9	29.7	27.4	29.3	35.0	30.5	36.6	28.8	21.8	15.3	29.9
VICT	40.0	28.7	29.3	33.8	36.8	29.2	30.8	28.1	30.5	36.0	30.5	37.1	29.9	26.6	15.7	30.9
(d) denoising SIDD-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	10.41	26.06	33.67	33.80	17.21	9.15	21.44	32.94	20.98	25.92	30.22	35.54	27.26	9.54	30.01	24.28
VICT	10.55	26.01	33.69	33.54	17.31	9.20	21.53	32.86	20.95	27.95	30.31	35.43	27.27	9.49	29.94	24.40
<i>one-shot setting:</i>																
Painter	10.91	25.83	33.38	33.83	16.94	9.54	22.65	33.17	22.37	25.64	30.12	35.56	27.02	9.92	29.97	24.46
VICT	13.95	26.06	33.02	34.87	15.93	12.07	24.85	33.31	23.88	28.74	30.86	35.52	27.43	11.89	30.25	25.51
(e) deraining Rain-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	12.37	14.63	20.13	21.07	14.30	12.98	22.35	20.30	22.27	26.23	19.00	23.91	21.32	13.54	18.07	18.83
VICT	12.49	14.70	20.16	21.24	14.35	13.11	22.85	20.38	22.75	26.27	19.05	24.06	21.72	13.82	18.11	19.00
<i>one-shot setting:</i>																
Painter	12.71	14.52	20.17	21.10	14.20	13.39	22.13	20.37	21.83	26.23	19.00	23.99	20.88	13.77	18.09	18.83
VICT	17.57	15.21	21.01	21.93	14.56	18.29	23.35	21.38	23.00	26.42	19.39	24.56	21.86	19.58	18.46	20.44
(f) low-light enhancement LoL-C (PSNR ↑)																
<i>zero-shot setting:</i>																
Painter	16.69	13.99	19.42	20.09	11.27	12.26	12.13	19.61	12.91	17.44	18.30	21.36	17.85	12.53	17.04	16.19
VICT	16.79	14.00	19.61	20.07	11.32	12.22	12.02	19.65	12.93	17.55	18.38	21.43	19.10	12.79	17.12	16.33
<i>one-shot setting:</i>																
Painter	17.01	14.11	19.43	20.08	11.30	12.46	12.67	19.63	13.77	17.40	18.40	21.37	17.35	11.90	17.02	16.26
VICT	17.82	16.05	19.47	20.14	11.38	12.44	12.81	19.54	15.32	17.59	18.59	21.26	19.87	12.86	17.20	16.82

References

- [1] Amir Bar, Yossi Gandelsman, Trevor Darrell, Amir Globerson, and Alexei Efros. Visual prompting via image inpainting. In *NeurIPS*, 2022. [1](#)
- [2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. [1](#)
- [3] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. [1](#)
- [4] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *CVPR*, 2017. [1](#)
- [5] Amir R Zamir, Alexander Sax, William Shen, Leonidas J Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *CVPR*, 2018. [1](#)