Exploring Resolution and Degradation Clues as Self-supervised Signal for Low Quality Object Detection

Ziteng Cui[1], Yingying Zhu[2], Lin Gu[3,1*], Guo-Jun Qi[4], Xiaoxiao Li[5], Renrui Zhang[6], Zenghui Zhang[7], Tatsuya Harada[1,3]

1. The University of Tokyo 2. University of Texas at Arlington 3. RIKEN AIP 4. Laboratory for Machine Perception and Learning
5. The University of British Columbia 6. Shanghai AI Laboratory 7. Shanghai Jiao Tong University

ECCV TEL AVIV 2022

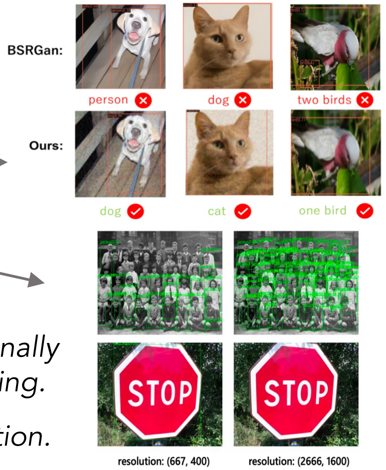
Motivation & Contribution:

Degradation Conditions:

1. *Noise, Blur, Low-Resolution Always Affect Vision Tasks.*
2. *Restoration methods may not such effectiveness.*
3. *Scale GAP in Object Detection Task.*

Our Contributions:

1. *Take the degradation types as self-supervised signals.*
2. *Combine super-resolution and object detection, we additionally design a ARRD decoder on detectors for self-supervised learning.*
3. *AERIS get SOTA performance, even with lower input resolution.*



Experiment Results (Best Results on both single and multi degradation):

Test Set	Pre-process	Training Strategy	CenterNet (ResNet-18) AP	AP_s	AP_m	AP_l	CenterNet (Swin-T) AP	AP_s	AP_m	AP_l
COCO	-	Detection	30.1	10.6	33.2	47.2	36.9	17.9	41.8	52.9
COCO-d	-	Detection	14.5	1.2	10.4	38.6	19.9	2.7	16.9	46.2
	bicubic (×2)		16.2	4.1	15.3	31.1	18.6	4.0	17.8	39.7
	bicubic (×4)		8.0	4.6	10.5	10.1	10.6	**5.7**	12.8	16.7
	SRGAN [31] (×2)		14.8	2.6	14.3	27.9	16.6	3.0	16.5	33.4
	DBPN [20] (×2)		15.0	3.5	14.3	27.4	16.7	3.4	16.1	32.0
	Real-SR [4] (×2)		14.2	2.6	12.4	29.5	17.3	3.6	17.0	34.1
	BSRGAN [60] (×2)		16.8	4.2	15.8	36.9	20.2	4.8	18.1	40.5
	BM3D		10.4	0.8	6.8	27.9	10.9	0.7	8.8	35.1
	Restormer [59]		11.4	1.2	7.2	34.8	11.9	1.4	8.9	33.4
	-	Deg t	17.6	2.3	15.4	41.9	20.9	3.1	20.3	47.6
		Deg t + N	17.9	2.5	15.9	42.5	21.0	3.0	20.4	48.2
		D_r + Detection	17.7	**4.8**	15.8	41.0	21.4	5.6	19.6	46.3
		AERIS	**18.4**	2.7	**16.4**	**42.5**	**21.6**	3.2	**20.4**	**49.0**

Multi-degradation condition, blue background means higher resolution.

(a) Noise.

Method \ σ	(5, 50)	15	25	50
-	22.8	26.8	23.8	15.4
IRCNN [62]	22.6	26.8	24.2	16.8
Swin-IR [33]	24.2	28.0	25.6	19.3
Restormer [59]	23.8	27.6	25.1	18.9
Deg t + N	24.3	27.6	25.0	18.3
D_r + Detection	24.8	28.5	25.5	19.4
AERIS	**25.1**	**28.7**	**26.5**	**20.2**

(b) Blur.

Method \ k	Mix			
-	26.7	25.8	23.9	23.1
EPLL [66]	27.8	26.8	25.4	25.2
IRCNN [62]	26.7	26.9	24.1	22.8
Deg t + N	28.8	27.5	27.6	27.8
D_r + Detection	28.5	27.7	27.5	27.3
AERIS	**29.3**	**28.6**	**28.0**	**28.2**

(c) Down-sampling (Ratio: **2**).

Method \ metric	AP	AP_s	AP_m	AP_l
-	20.2	1.5	16.1	49.8
SRGAN [31] (×2)	24.0	6.2	25.4	39.8
DBPN [20] (×2)	25.1	7.3	27.1	41.9
Swin-IR [33] (×2)	25.4	7.6	27.0	42.4
D_r + Detection	26.0	8.4	26.2	46.5
AERIS	25.4	4.6	25.8	**49.6**
AERIS (×2)	**26.8**	**8.6**	**28.8**	45.2

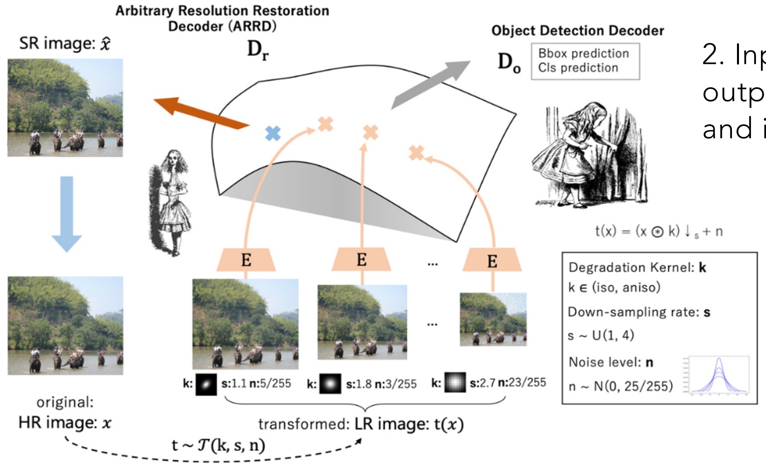
(d) Down-sampling (Ratio: **4**).

Method \ metric	AP	AP_s	AP_m	AP_l
-	8.2	0.0	3.2	33.1
DBPN [20] (×2)	14.8	1.0	9.9	39.7
DBPN [20] (×4)	12.2	1.7	11.7	23.4
Swin-IR [33] (×2)	15.2	1.1	10.1	39.9
Swin-IR [33] (×4)	12.8	1.8	12.2	23.4
D_r + Detection	15.1	1.8	12.7	40.1
AERIS	13.0	0.8	10.2	**42.6**
AERIS (×2)	**15.8**	**2.0**	**13.2**	40.9

Single-degradation condition, noise/ blur/ low-resolution.

Demo Show:
(a). Original Image
(b) Low-Resolution Degraded Image
(c)/(d)/(e): Restoration Methods
(f) Our AERIS
[background is outputs of ARRD]

Proposed Method (AERIS, Auto-Encoding Resolution In Self-supervision):



2. Input the t(x) to encoder E, and output both for $\boxed{\text{object detection}}$ and image restoration:

self-supervised learning	supervised learning
decoder: Dd	decoder: Do

Arbitrary Resolution Restoration Decoder (ARRD)

1. x ----> t(x): Base on the image degradation function $t(x) = (x \circledast k) \downarrow_s + n$,

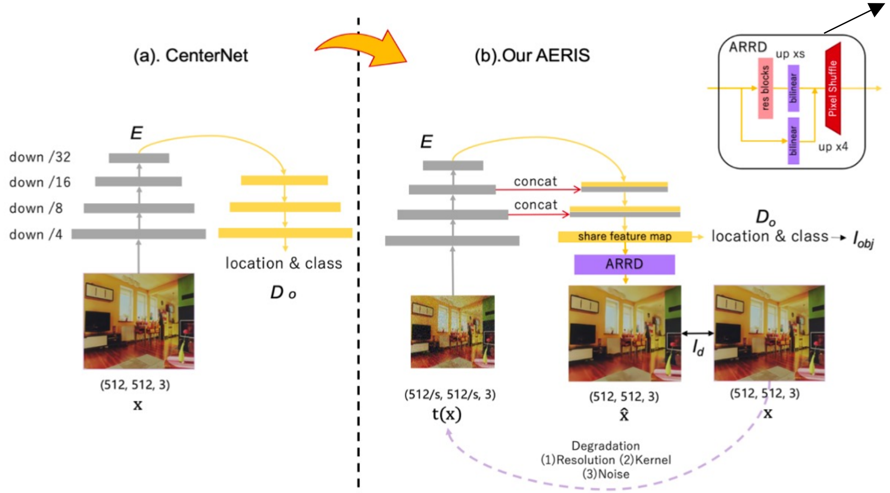
Blur Kernel k :
isotropic Gaussian kernels k_{iso}
anisotropic Gaussian kernels k_{aniso}

Noise n :
Zero-mean additive white Gaussian noise
$n \sim N(0, \sigma)$ $\sigma \sim U(0, 25/255)$ (*e.g.* 13.2/255).

Resolution s :
Random from 1~4
Type: Bicubic/ Bilinear/ Nearest

Implement on the detector (CenterNet for example):

ARRD: residual bilinear with pixel-shuffle



Restoration loss:

$$l_d = |\hat{x} - x|_1 = |D_r(E(t(x))) - x|_1.$$

Total loss: object detection loss

$$l_{total} = l_{obj} + \lambda \cdot l_d,$$