Appendix

In this appendix, we will add more details about some experiments reported in the paper, including

1. The accuracy of classification experiment on ImageNet-C of each corruption, average across all corruption levels, and with corruption level 1-5 respestively.

2. The precision recall curve for retrieval task on ImageNet and ImageNet-C of each corruption with corruption level 1-5.

3. The precision recall curve for retrieval task on CIFAR-10 and CIFAR-10-C of each corruption with corruption level 1-5.

A. Classification accuracy on ImageNet-C

In Table Supp.1, we compare the classification performance on ImageNet-C for baseline contrastive model and the proposed models trained with spatial-attention teacher guidance, average over the 5 different corruption/noise magnitudes in ImageNet-C. We further show the classification accuracy on ImageNet-C data set of each corruption type with corruption level 1-5 in Table Supp.2 to Supp.3 here, for our proposed model vs baseline. We can see that the proposed model outperforms baseline for most cases.

B. Retrieval results on ImageNet and ImageNet-C

In Fig Supp.1, we show the retrieval results on ImageNet and ImageNet-C (and CIFAR-10 and CIFAR-10-C) with one corruption: Fog.

In Fig Supp.2 to Fig Supp.6, we show the retrieval results on ImageNet ("clean"), and ImageNet-C ("noise level 1-5") of every type of corruption/noise, for the proposed method ("Contrastive attn teacher") compared to the baseline SimCLR ("Contrastive"). We can see that the proposed method obtains better results than the baseline for most corruption/noise types, except "contrast" corruption only.

C. Retrieval results on CIFAR-10 and CIFAR-10-C

In Fig Supp.7 to Fig Supp.11, we show the retrieval results on CIFAR-10 ("clean"), and CIFAR-10-C ("noise level 1-5") of every type of corruption/noise, for the proposed method ("Contrastive attn teacher") compared to the baseline SimCLR ("Contrastive"). We can see that the proposed method outperformed the baseline for all corruption/noise types.

D. Supplementary tables

Table Supp.1. ImageNet-C Top-1 classification accuracy for different models (mean $\pm$ SE for 3 seeds).

| Model	Brightness	Contrast	Elastic	Pixelate	JPEG
Contrastive	60.10 ± 0.22	52.07 ± 1.24	33.00 ± 0.31	46.49 ± 0.36	45.78 ± 0.36
Contrastive attn. teacher	$\mathbf{60.94} \pm \mathbf{0.15}$	$\mathbf{52.18} \pm \mathbf{0.16}$	$\mathbf{34.42} \pm \mathbf{0.33}$	$\mathbf{48.27} \pm \mathbf{1.46}$	$\mathbf{46.41} \pm \mathbf{0.60}$
Model	Gaussian	Shot	Impluse	Defocus	Glass
Contrastive	28.15 ± 0.27	24.18 ± 0.41	$\mathbf{16.64} \pm \mathbf{1.23}$	23.61 ± 0.47	13.45 ± 0.68
Contrastive attn. teacher	$\mathbf{28.30} \pm \mathbf{0.54}$	$\mathbf{24.84} \pm \mathbf{0.26}$	16.49 ± 0.89	$\mathbf{24.45} \pm \mathbf{0.59}$	$\mathbf{13.87} \pm \mathbf{0.08}$
Model	Motion	Zoom	Snow	Frost	Fog
Contrastive	22.21 ± 0.45	18.95 ± 0.60	28.21 ± 0.08	29.86 ± 0.18	45.34 ± 0.53
Contrastive attn. teacher	$\mathbf{22.81} \pm \mathbf{1.06}$	$\mathbf{19.10} \pm \mathbf{0.23}$	$\mathbf{29.78} \pm \mathbf{0.21}$	$\mathbf{31.13} \pm \mathbf{0.63}$	$\mathbf{46.12} \pm \mathbf{0.52}$
Model	Speckle Noise	Gaussian Blur	Spatter	Saturate	Average
Contrastive	28.23 ± 0.31	26.16 ± 0.07	43.08 ± 0.18	60.42 ± 0.15	34.00 ± 0.06
Contrastive attn. teacher	$\mathbf{29.15} \pm \mathbf{0.65}$	$\mathbf{27.10} \pm \mathbf{0.35}$	$\mathbf{44.04} \pm \mathbf{0.08}$	$\mathbf{60.50} \pm \mathbf{0.02}$	$\mathbf{34.73} \pm \mathbf{0.06}$

Table Supp.2. Classification accruacy on ImageNet-C for each corruption type with corruption level 1-5.

Model	Brightness 1	Brightness 2	Brightness 3	Brightness 4	Brightness 5
Contrastive	65.32 ± 0.04	63.92 ± 0.12	61.57 ± 0.16	57.67 ± 0.26	52.05 ± 0.55
Contrastive attn. teacher	$\mathbf{66.07} \pm \mathbf{0.13}$	$\mathbf{64.78} \pm \mathbf{0.22}$	$\mathbf{62.42} \pm \mathbf{0.09}$	$\mathbf{58.51} \pm \mathbf{0.19}$	$\mathbf{52.96} \pm \mathbf{0.26}$

Model	Contrast 1	Contrast 2	Contrast 3	Contrast 4	Contrast 5
Contrastive	64.21 ± 0.14	63.46 ± 0.26	61.54 ± 0.56	$\mathbf{50.55} \pm \mathbf{1.74}$	20.60 ± 3.61
Contrastive attn. teacher	$\mathbf{64.78} \pm \mathbf{0.05}$	$\mathbf{63.97} \pm \mathbf{0.03}$	$\mathbf{61.79} \pm \mathbf{0.07}$	49.66 ± 0.53	$\mathbf{20.62} \pm \mathbf{0.92}$

Model	Elastic 1	Elastic 2	Elastic 3	Elastic 4	Elastic 5
Contrastive	57.53 ± 0.13	36.38 ± 0.10	40.44 ± 0.56	24.60 ± 0.76	6.03 ± 0.30
Contrastive attn. teacher	$\mathbf{58.40} \pm \mathbf{0.35}$	$\mathbf{37.76} \pm \mathbf{0.32}$	$\mathbf{42.40} \pm \mathbf{0.38}$	$\mathbf{26.78} \pm \mathbf{0.42}$	$\mathbf{6.75} \pm \mathbf{0.40}$

Model	JPEG 1	JPEG 2	JPEG 3	JPEG 4	JPEG 5
Contrastive	54.74 ± 0.64	51.52 ± 0.81	49.02 ± 0.86	41.55 ± 0.72	32.07 ± 0.62
Contrastive attn. teacher	$\mathbf{55.79} \pm \mathbf{0.33}$	$\mathbf{52.33} \pm \mathbf{0.47}$	$\mathbf{49.78} \pm \mathbf{0.58}$	$\mathbf{42.01} \pm \mathbf{0.66}$	$\mathbf{32.14} \pm \mathbf{1.06}$

Model	Gaussian 1	Gaussian 2	Gaussian 3	Gaussian 4	Gaussian 5
Contrastive	$\mathbf{53.51} \pm \mathbf{0.39}$	$\mathbf{43.37} \pm \mathbf{0.52}$	$\mathbf{27.95} \pm \mathbf{0.35}$	12.64 ± 0.81	3.30 ± 0.58
Contrastive attn. teacher	53.37 ± 0.30	43.24 ± 0.54	27.93 ± 0.71	$\mathbf{13.15} \pm \mathbf{0.81}$	$\mathbf{3.63} \pm \mathbf{0.51}$

Model	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5
Contrastive	49.72 ± 0.42	36.65 ± 0.12	22.84 ± 0.54	8.17 ± 0.86	3.51 ± 0.61
Contrastive attn. teacher	$\mathbf{50.51} \pm \mathbf{0.53}$	$\mathbf{37.49} \pm \mathbf{0.53}$	$\mathbf{23.58} \pm \mathbf{0.21}$	$\mathbf{8.82} \pm \mathbf{0.69}$	$\mathbf{3.79} \pm \mathbf{0.71}$

Model	Impulse 1	Impulse 2	Impulse 3	Impulse 4	Impulse 5
Contrastive	$\mathbf{36.02} \pm \mathbf{2.22}$	$\mathbf{23.52} \pm \mathbf{1.88}$	$\mathbf{15.49} \pm \mathbf{1.38}$	5.89 ± 0.54	2.26 ± 0.28
Contrastive attn. teacher	35.00 ± 1.99	22.83 ± 1.77	15.39 ± 0.84	$\mathbf{6.61} \pm \mathbf{0.26}$	$\mathbf{2.63} \pm \mathbf{0.34}$

Model	Pixelate 1	Pixelate 2	Pixelate 3	Pixelate 4	Pixelate 5
Contrastive	57.80 ± 0.19	54.04 ± 0.38	48.53 ± 0.56	39.13 ± 0.43	32.94 ± 0.30
Contrastive attn. teacher	$\mathbf{58.49} \pm \mathbf{1.14}$	$\mathbf{55.05} \pm \mathbf{1.25}$	$\mathbf{50.37} \pm \mathbf{1.63}$	$\mathbf{41.64} \pm \mathbf{1.82}$	$\mathbf{35.81} \pm \mathbf{1.54}$

Model	Defocus 1	Defocus 2	Defocus 3	Defocus 4	Defocus 5
Contrastive	43.46 ± 0.59	34.78 ± 0.91	20.78 ± 0.70	11.95 ± 0.35	7.06 ± 0.49
Contrastive attn. teacher	$\mathbf{44.07} \pm \mathbf{0.17}$	$\mathbf{35.43} \pm \mathbf{0.42}$	$\mathbf{21.84} \pm \mathbf{0.82}$	$\mathbf{13.14} \pm \mathbf{0.95}$	$\mathbf{7.80} \pm \mathbf{0.78}$

Model	Glass 1	Glass 2	Glass 3	Glass 4	Glass 5
Contrastive	36.65 ± 1.10	20.88 ± 1.22	4.97 ± 0.55	$\mathbf{2.92} \pm \mathbf{0.34}$	1.85 ± 0.17
Contrastive attn. teacher	$\mathbf{37.71} \pm \mathbf{0.60}$	$\mathbf{21.79} \pm \mathbf{0.17}$	$\mathbf{5.17} \pm \mathbf{0.40}$	2.80 ± 0.17	$\mathbf{1.89} \pm \mathbf{0.10}$

Table Supp.3. Classification accruacy on ImageNet-C for each corruption type with corruption level 1-5.

Model	Motion 1	Motion 2	Motion 3	Motion 4	Motion 5
Contrastive	52.49 ± 0.53	35.65 ± 0.73	14.49 ± 0.76	5.30 ± 0.53	3.15 ± 0.36
Contrastive attn. teacher	$\mathbf{53.33} \pm \mathbf{0.24}$	$\mathbf{36.79} \pm \mathbf{1.87}$	$\mathbf{15.16} \pm \mathbf{2.31}$	$\mathbf{5.53} \pm \mathbf{0.65}$	$\mathbf{3.23} \pm \mathbf{0.28}$

Model	Zoom 1	Zoom 2	Zoom 3	Zoom 4	Zoom 5
Contrastive	31.43 ± 0.57	22.00 ± 0.57	18.18 ± 0.75	13.23 ± 0.67	9.89 ± 0.57
Contrastive attn. teacher	$\mathbf{31.83} \pm \mathbf{0.31}$	$\mathbf{22.23} \pm \mathbf{0.15}$	$\mathbf{18.22} \pm \mathbf{0.39}$	$\mathbf{13.25} \pm \mathbf{0.47}$	$\mathbf{9.99} \pm \mathbf{0.37}$

Model	Snow 1	Snow 2	Snow 3	Snow 4	Snow 5
Contrastive	45.62 ± 0.24	25.89 ± 0.14	30.01 ± 0.29	21.94 ± 0.25	17.58 ± 0.24
Contrastive attn. teacher	$\mathbf{47.07} \pm \mathbf{0.29}$	$\mathbf{27.78} \pm \mathbf{0.25}$	$\mathbf{31.54} \pm \mathbf{0.36}$	$\mathbf{23.46} \pm \mathbf{0.44}$	$\mathbf{19.08} \pm \mathbf{0.22}$

Model	Frost 1	Frost 2	Frost 3	Frost 4	Frost 5
Contrastive	51.32 ± 0.28	34.38 ± 0.13	24.29 ± 0.23	22.14 ± 0.14	17.20 ± 0.15
Contrastive attn. teacher	$\mathbf{52.43} \pm \mathbf{0.32}$	$\mathbf{35.76} \pm \mathbf{0.69}$	$\mathbf{25.65} \pm \mathbf{0.81}$	$\mathbf{23.53} \pm \mathbf{0.71}$	$\mathbf{18.29} \pm \mathbf{0.65}$

Model	Fog 1	Fog 2	Fog 3	Fog 4	Fog 5
Contrastive	56.81 ± 0.37	52.51 ± 0.45	45.87 ± 0.61	41.79 ± 0.62	29.71 ± 0.62
Contrastive attn. teacher	$\mathbf{57.39} \pm \mathbf{0.29}$	$\mathbf{53.03} \pm \mathbf{0.54}$	$\mathbf{46.68} \pm \mathbf{0.48}$	$\mathbf{42.72} \pm \mathbf{0.61}$	$\mathbf{30.79} \pm \mathbf{0.69}$

Model	Speckle Noise 1	Speckle Noise 2	Speckle Noise 3	Speckle Noise 4	Speckle Noise 5
Contrastive	50.53 ± 0.52	42.83 ± 0.43	23.33 ± 0.24	15.47 ± 0.21	8.99 ± 0.40
Contrastive attn. teacher	$\mathbf{51.48} \pm \mathbf{0.19}$	$\mathbf{43.89} \pm \mathbf{0.64}$	$\mathbf{24.49} \pm \mathbf{1.18}$	$\mathbf{16.25} \pm \mathbf{0.89}$	$\mathbf{9.61} \pm \mathbf{0.93}$

Model	Gaussian Blur 1	Gaussian Blur 2	Gaussian Blur 3	Gaussian Blur 4	Gaussian Blur 5
Contrastive	54.82 ± 0.44	37.38 ± 0.60	22.42 ± 0.09	12.36 ± 0.51	3.81 ± 0.40
Contrastive attn. teacher	$\mathbf{55.97} \pm \mathbf{0.77}$	$\mathbf{37.92} \pm \mathbf{0.22}$	$\mathbf{23.37} \pm \mathbf{0.86}$	$\mathbf{13.74} \pm \mathbf{1.03}$	$\mathbf{4.49} \pm \mathbf{0.52}$

Model	Spatter 1	Spatter 2	Spatter 3	Spatter 4	Spatter 5
Contrastive	63.95 ± 0.30	53.99 ± 0.15	41.75 ± 0.12	33.13 ± 0.37	22.59 ± 0.23
Contrastive attn. teacher	$\mathbf{64.70} \pm \mathbf{0.17}$	$\mathbf{54.77} \pm \mathbf{0.15}$	$\mathbf{42.66} \pm \mathbf{0.41}$	$\mathbf{34.37} \pm \mathbf{0.39}$	$\mathbf{23.68} \pm \mathbf{0.36}$

Model	Saturate 1	Saturate 2	Saturate 3	Saturate 4	Saturate 5
Contrastive	64.10 ± 0.24	$\mathbf{61.33} \pm \mathbf{0.09}$	64.85 ± 0.17	59.98 ± 0.20	$\mathbf{51.83} \pm \mathbf{0.21}$
Contrastive attn. teacher	$\mathbf{64.35} \pm \mathbf{0.27}$	60.63 ± 0.27	$\mathbf{65.45} \pm \mathbf{0.16}$	$\mathbf{60.42} \pm \mathbf{0.20}$	51.64 ± 0.32

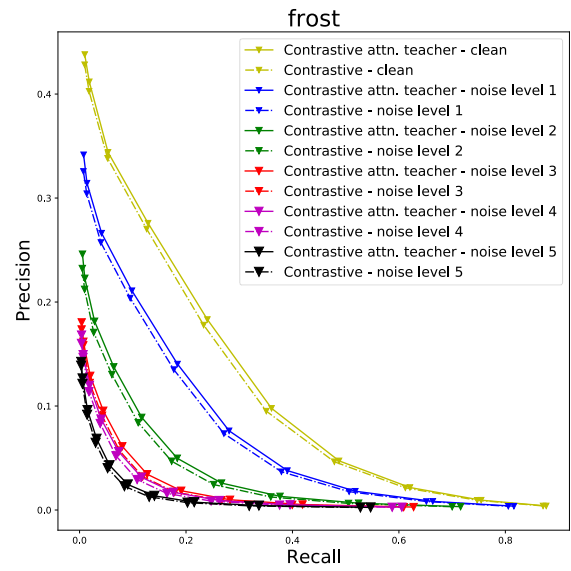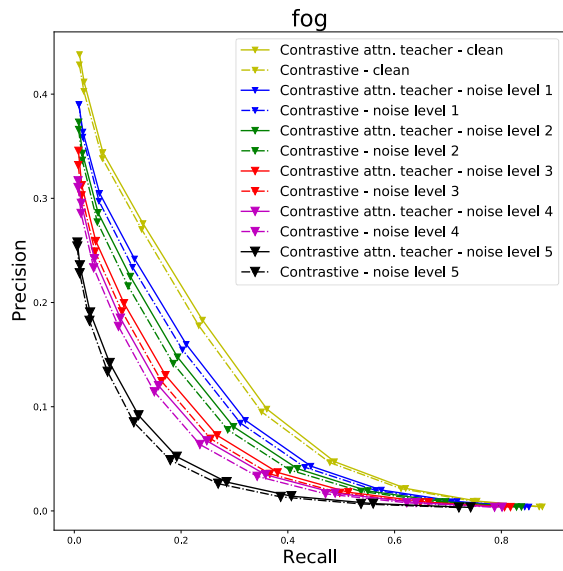# E. Supplementary figures
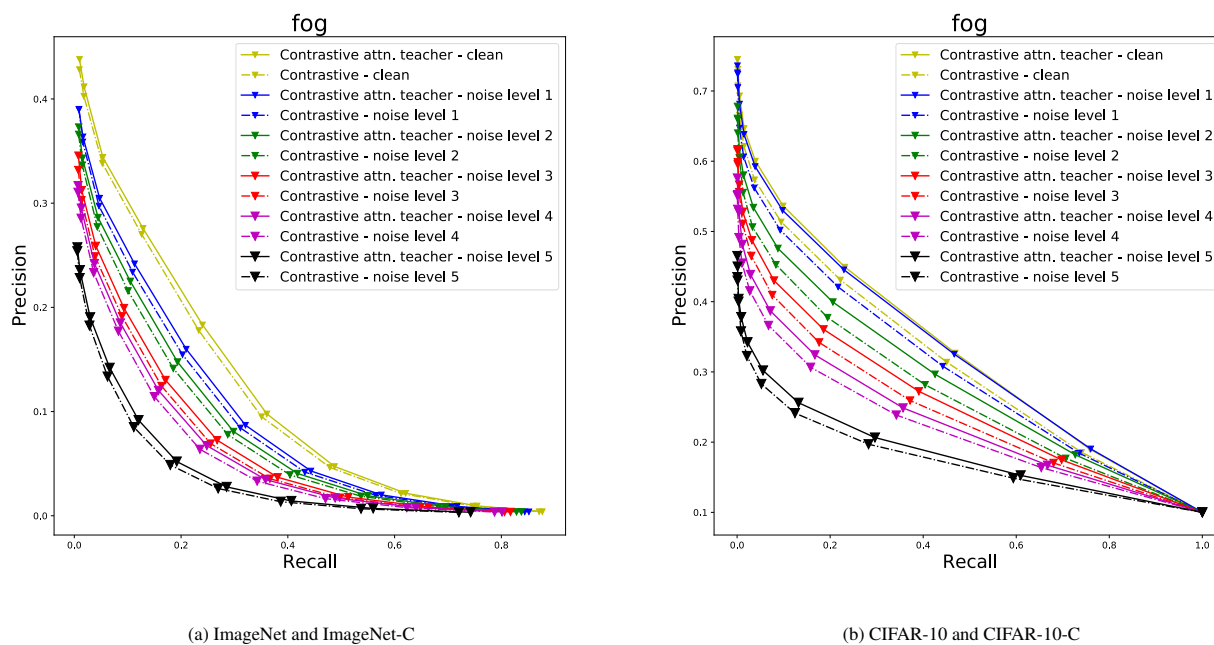
(a) ImageNet and ImageNet-C

(b) CIFAR-10 and CIFAR-10-C

Figure Supp.1. (a) Precision recall curve for retrieval on ImageNet validation set ("clean") and ImageNet-C with fog corruption ("noise level 1-5"). (b) CIFAR-10 test set and CIFAR10-C. Figures for other corruption types on both data sets can be found in the supplementary material.
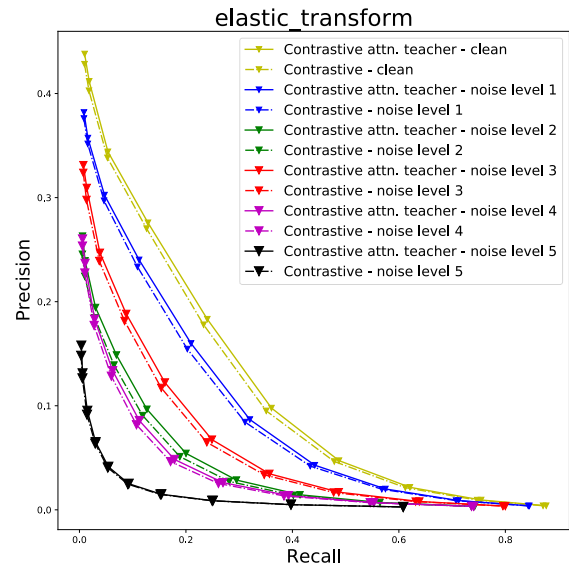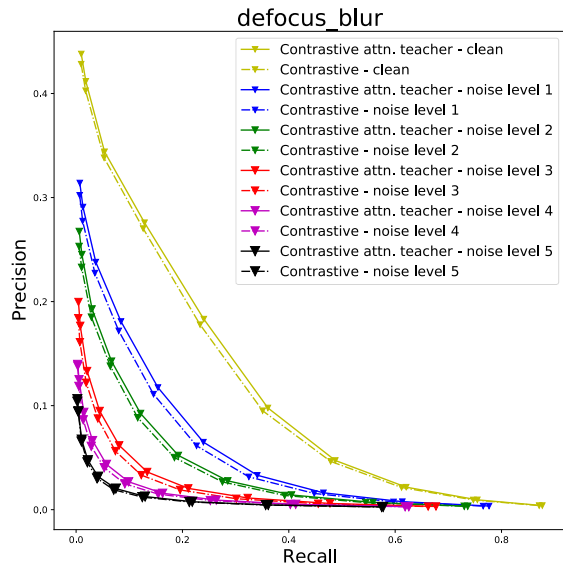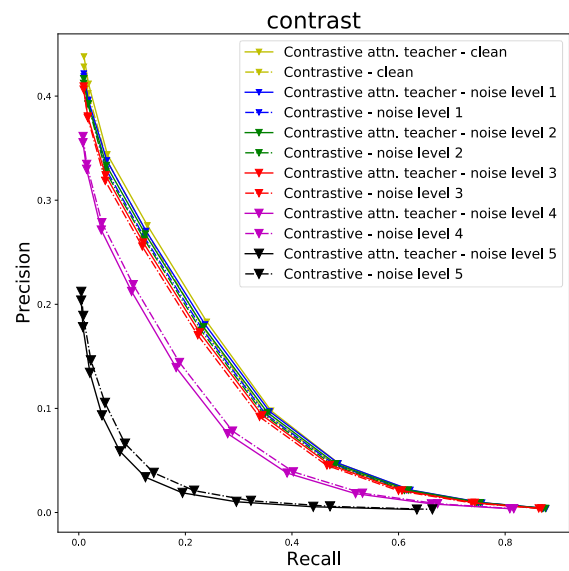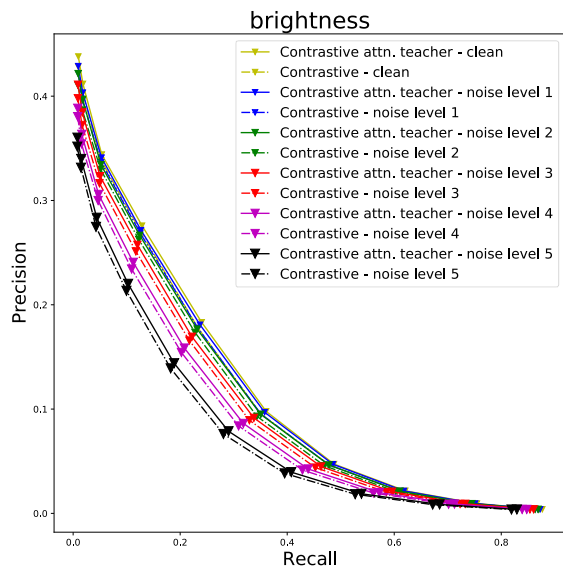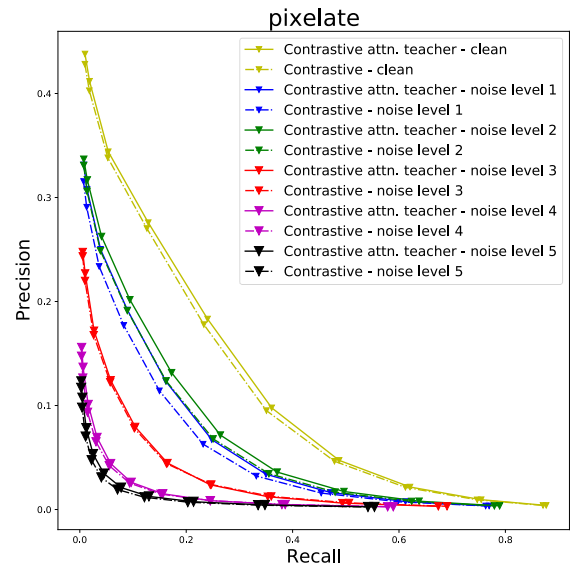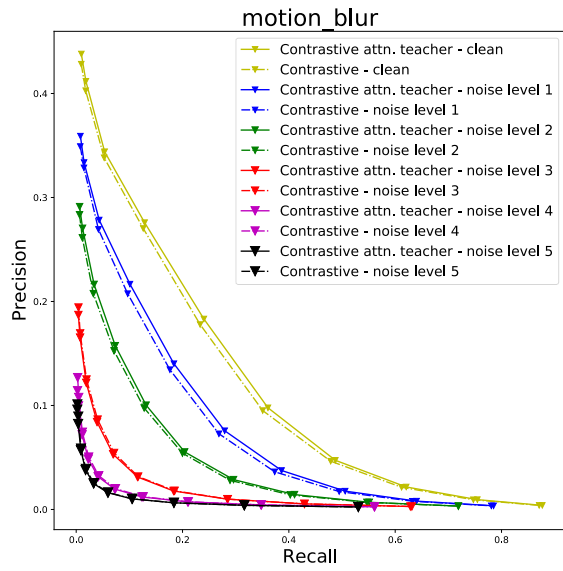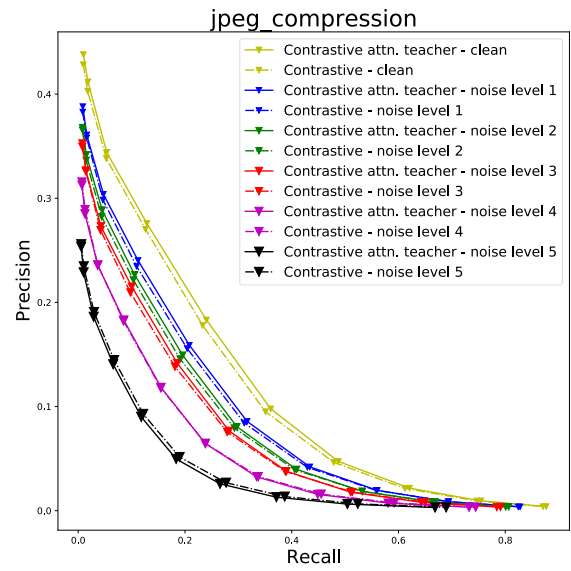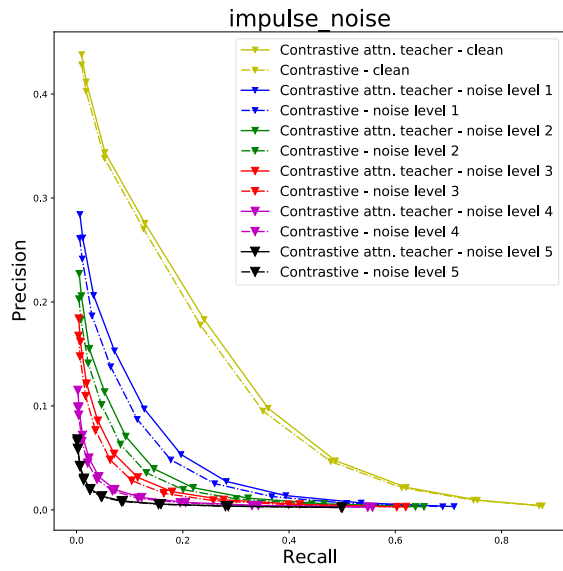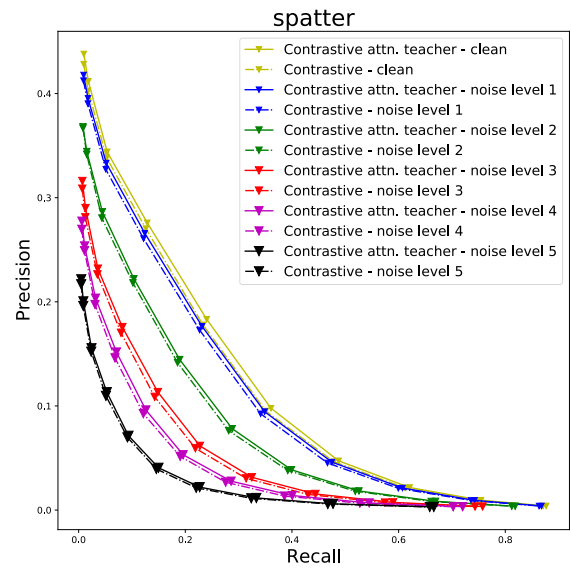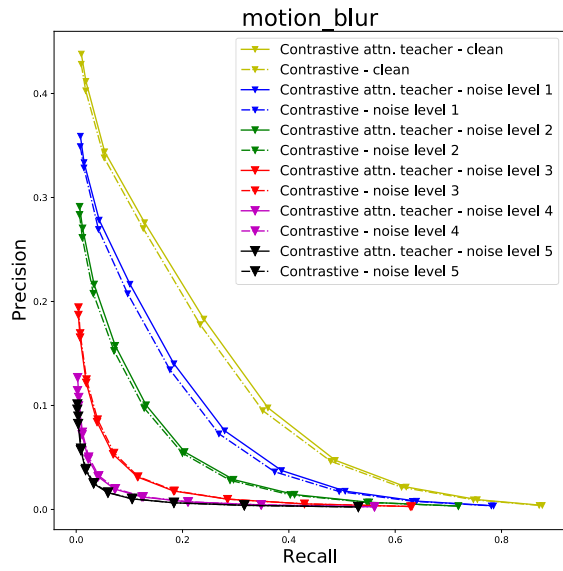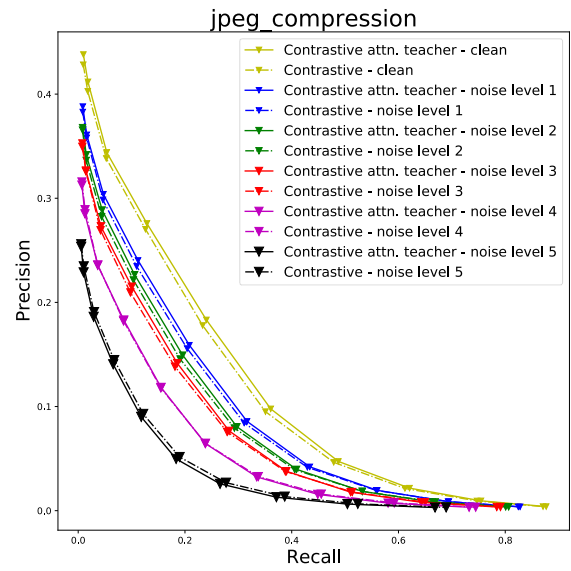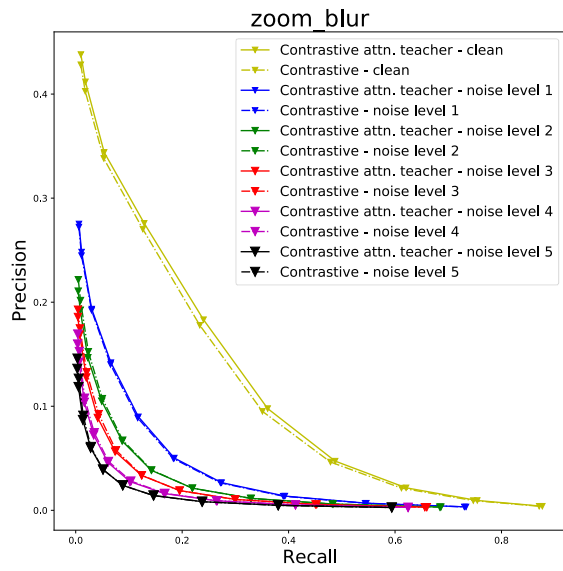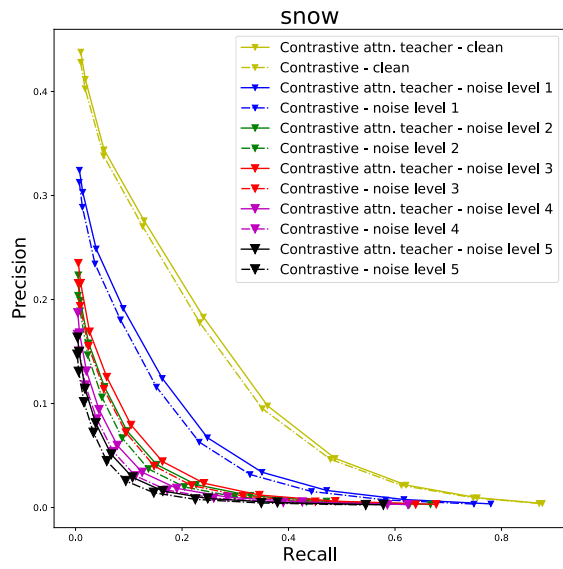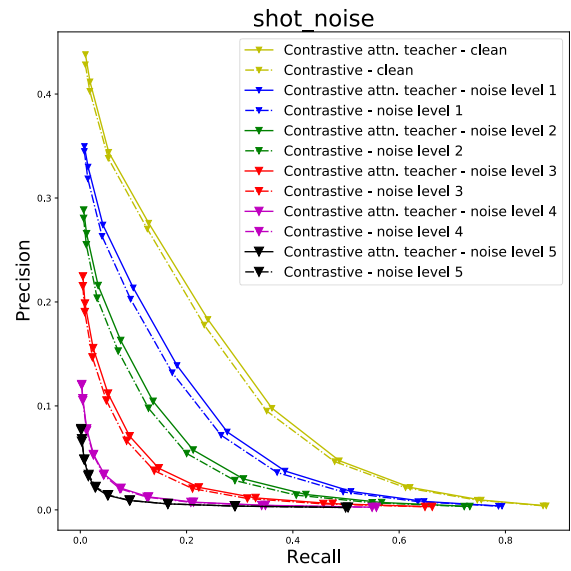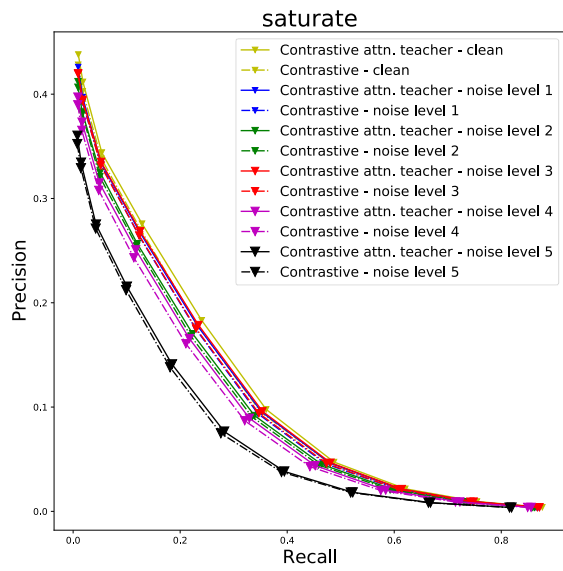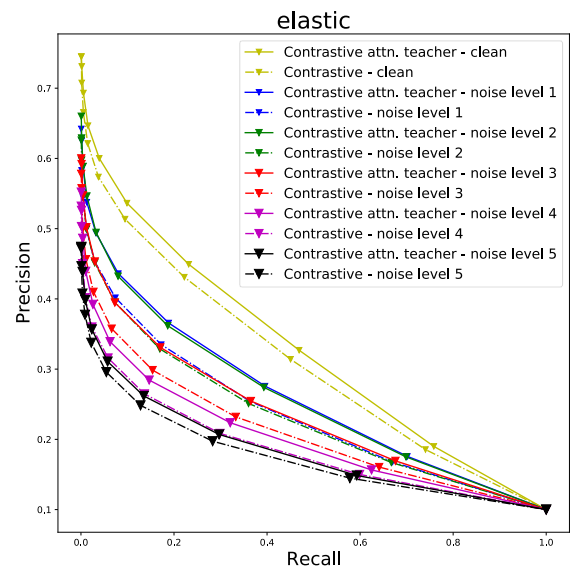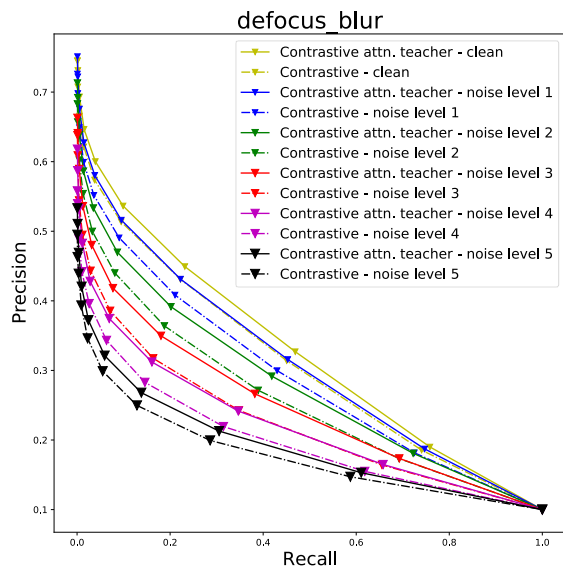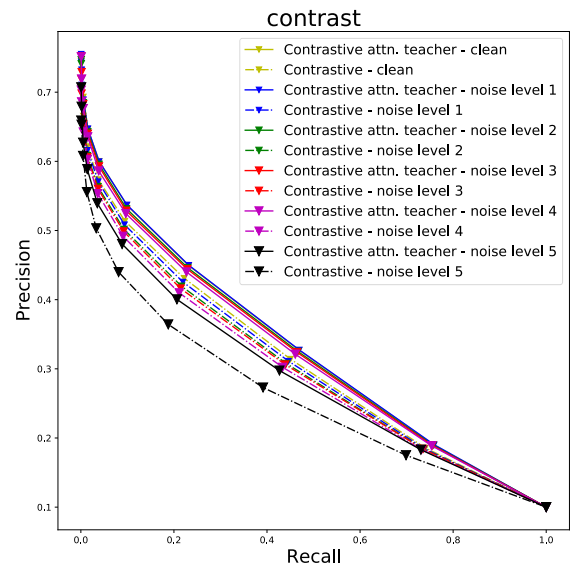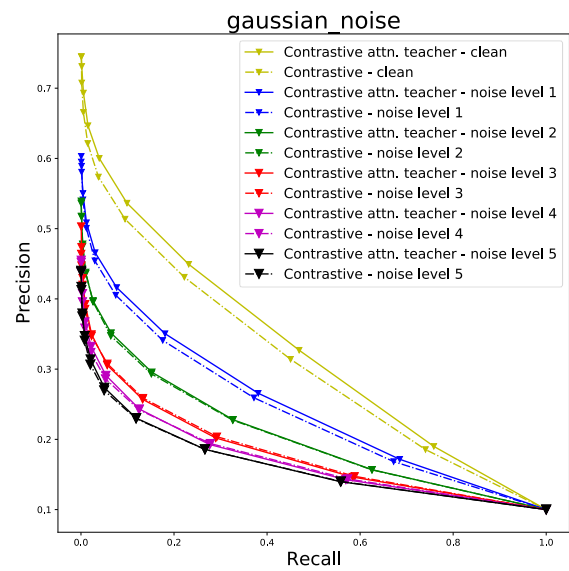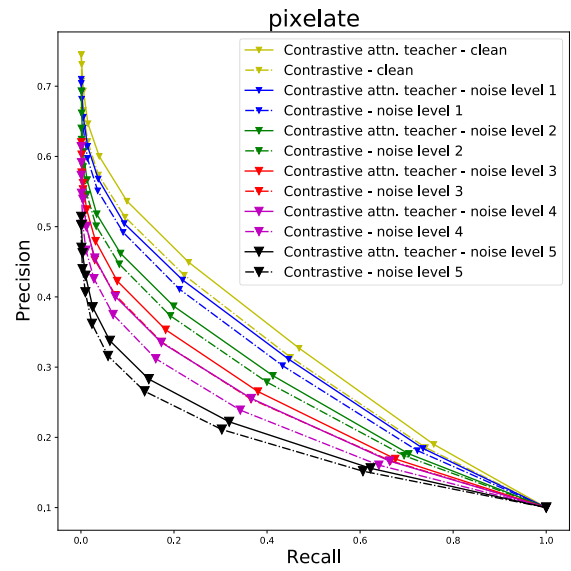
Figure Supp.2. Precision recall curve for retrieval on ImageNet validation set ("clean") and ImageNet-C with 4 corruptions, each with 5 corruption/noise level ("noise level 1-5").
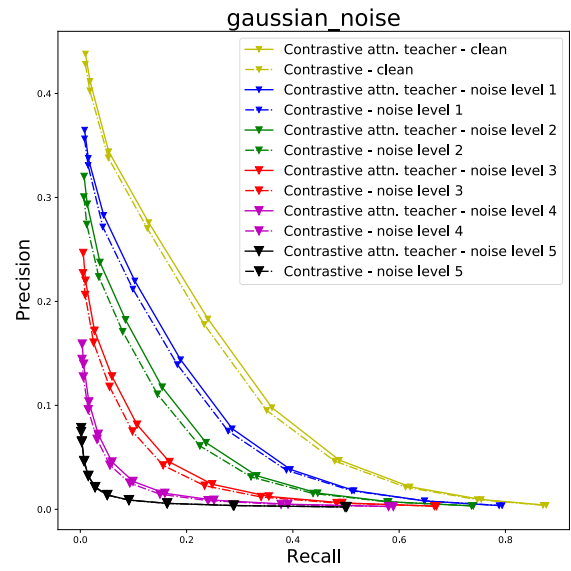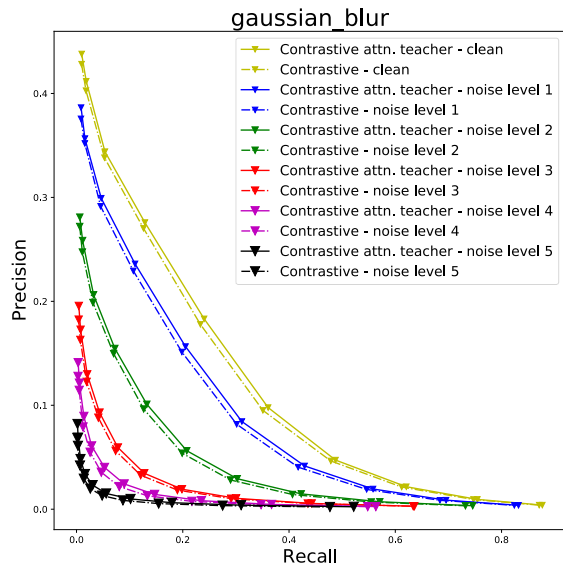
Figure Supp.3. Precision recall curve for retrieval on ImageNet validation set ("clean") and ImageNet-C with 4 corruptions, each with 5 corruption/noise level ("noise level 1-5").

Figure Supp.4. Precision recall curve for retrieval on ImageNet validation set ("clean") and ImageNet-C with 4 corruptions, each with 5 corruption/noise level ("noise level 1-5").

Figure Supp.5. Precision recall curve for retrieval on ImageNet validation set ("clean") and ImageNet-C with 4 corruptions, each with 5 corruption/noise level ("noise level 1-5").

Figure Supp.6. Precision recall curve for retrieval on ImageNet validation set ("clean") and ImageNet-C with 4 corruptions, each with 5 corruption/noise level ("noise level 1-5").
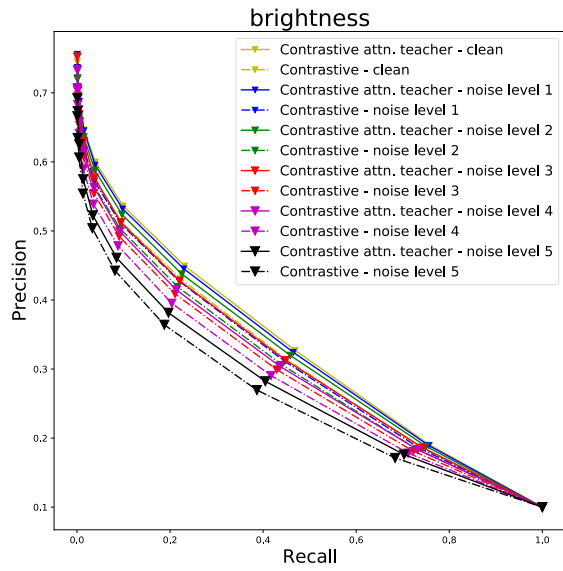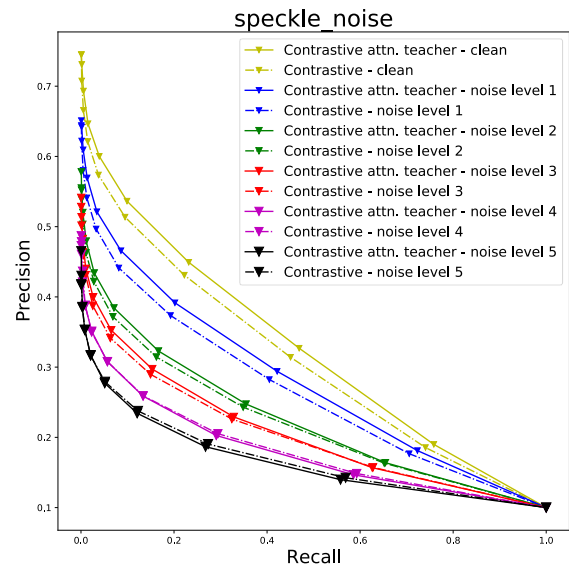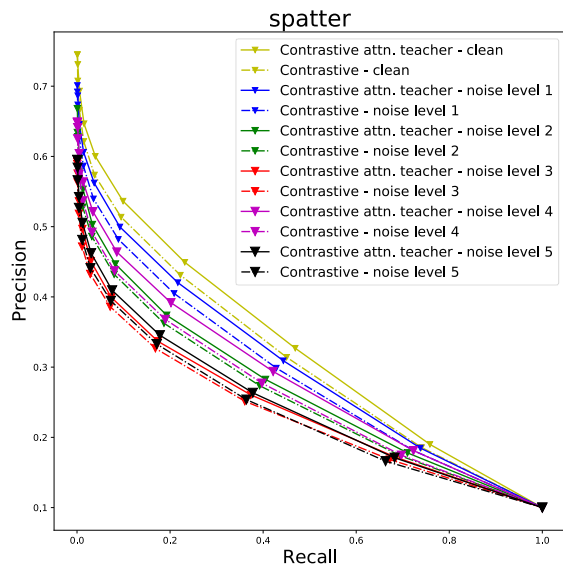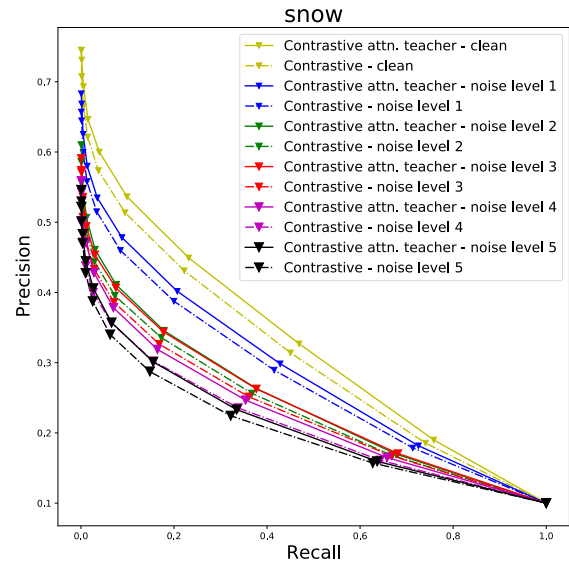
Figure Supp.7. Precision recall curve for retrieval on CIFAR-10 test set ("clean") and CIFAR10-C ("noise level 1-5").
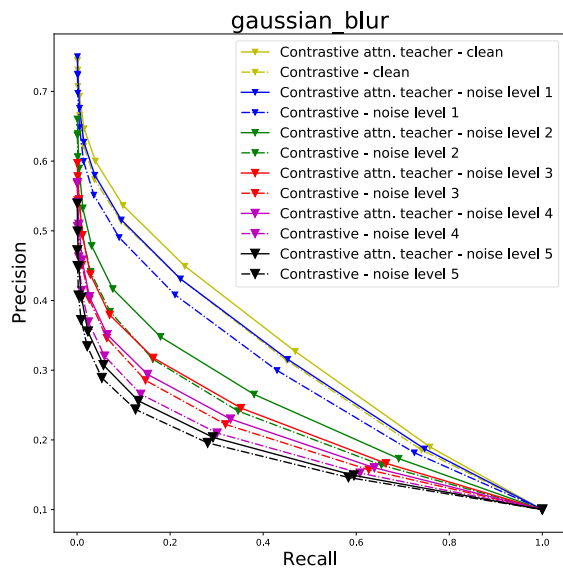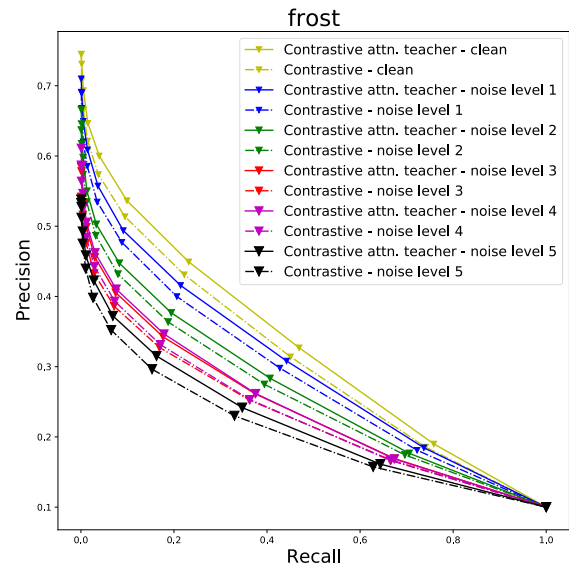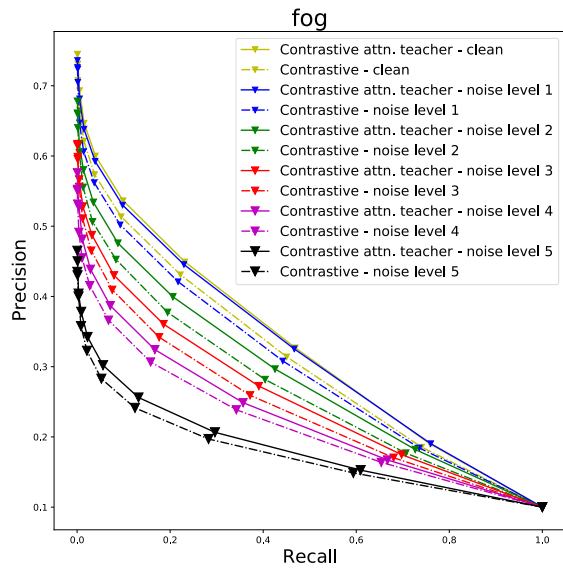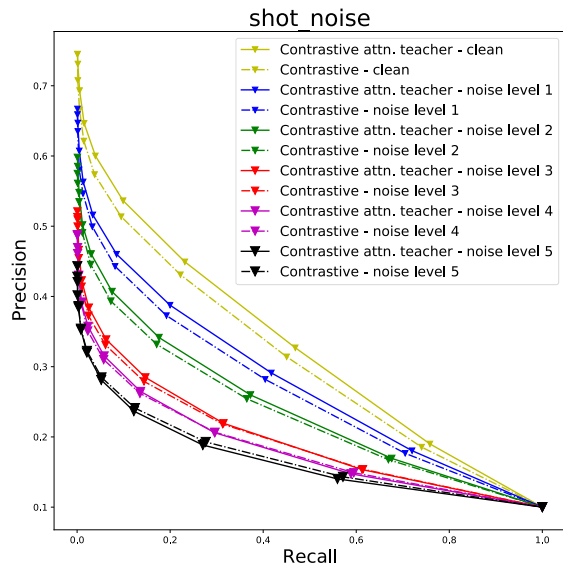
Figure Supp.8. Precision recall curve for retrieval on CIFAR-10 test set ("clean") and CIFAR10-C ("noise level 1-5").
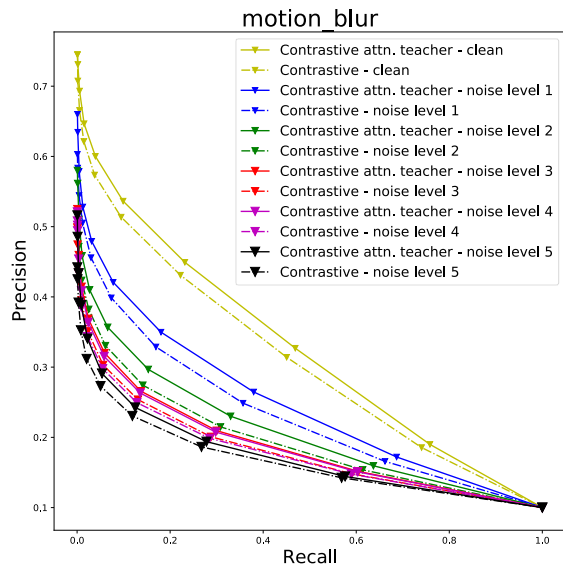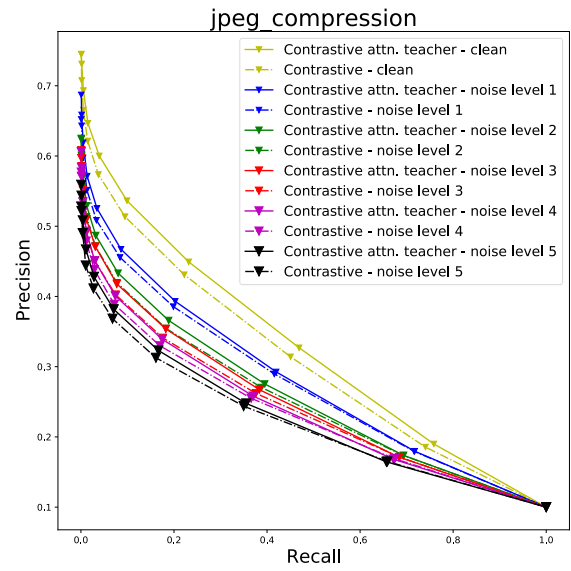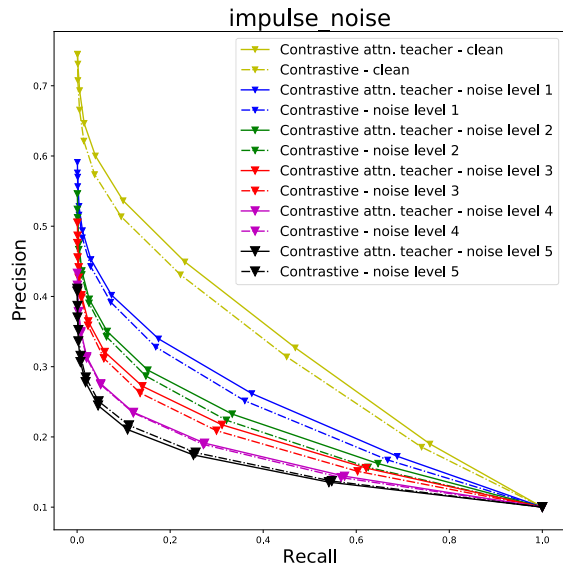
Figure Supp.9. Precision recall curve for retrieval on CIFAR-10 test set ("clean") and CIFAR10-C ("noise level 1-5").

Figure Supp.10. Precision recall curve for retrieval on CIFAR-10 test set ("clean") and CIFAR10-C ("noise level 1-5").
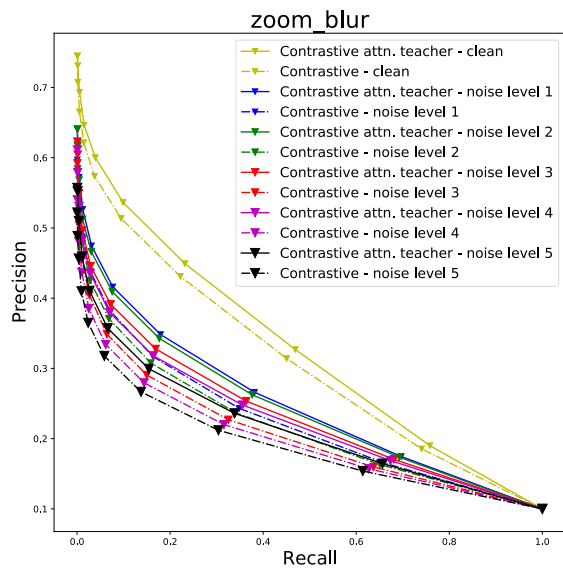
Figure Supp.11. Precision recall curve for retrieval on CIFAR-10 test set ("clean") and CIFAR10-C ("noise level 1-5").