

TABLE I: Comparison robustness performances among different data corruption in the lung segmentation task on **MC** dataset using UNet as baseline. Dice is utilized as the evaluation metric.

Method	UNet (—)	UNet (+ VAE)	UNet (+Ours)	Rec-UNet (—)	Rec-UNet (+Ours)
Original	$89.95 \pm 0.3$	$92.80 \pm 0.3$	$95.35 \pm 0.2$	$92.97 \pm 0.3$	<b><math>95.89 \pm 0.1</math></b>
Gauss. Noise	$71.32 \pm 0.2$	$90.87 \pm 0.3$	$89.66 \pm 0.2$	$81.67 \pm 0.1$	<b><math>93.47 \pm 0.2</math></b>
Shot Noise	$86.86 \pm 0.1$	$89.04 \pm 0.2$	$94.20 \pm 0.2$	$86.82 \pm 0.2$	<b><math>94.33 \pm 0.3</math></b>
Impulse Noise	$69.00 \pm 0.3$	$90.23 \pm 0.2$	<b><math>93.95 \pm 0.2</math></b>	$78.13 \pm 0.4$	$93.40 \pm 0.2$
Speckle Noise	$88.39 \pm 0.2$	$90.45 \pm 0.4$	$94.17 \pm 0.4$	$87.72 \pm 0.2$	<b><math>94.78 \pm 0.1</math></b>
Poisson Noise	$71.66 \pm 0.4$	$89.94 \pm 0.2$	<b><math>90.94 \pm 0.4</math></b>	$75.54 \pm 0.2$	$87.23 \pm 0.4$
Dropout	$87.62 \pm 0.3$	$92.31 \pm 0.1$	$93.24 \pm 0.2$	$89.81 \pm 0.1$	<b><math>94.10 \pm 0.1</math></b>
Gauss. Blur	$90.44 \pm 0.4$	$89.08 \pm 0.4$	$94.83 \pm 0.2$	$84.39 \pm 0.3$	<b><math>95.55 \pm 0.3</math></b>
Glass Blur	$88.45 \pm 0.3$	$89.81 \pm 0.2$	$94.45 \pm 0.3$	$91.92 \pm 0.3$	<b><math>95.28 \pm 0.4</math></b>
Defocus Blur	$89.79 \pm 0.3$	$89.56 \pm 0.3$	$94.75 \pm 0.3$	$88.35 \pm 0.2$	<b><math>95.25 \pm 0.5</math></b>
Motion Blur	$86.62 \pm 0.3$	$87.53 \pm 0.1$	<b><math>90.17 \pm 0.5</math></b>	$84.31 \pm 0.3$	$90.10 \pm 0.2$
Zoom Blur	$82.97 \pm 0.4$	$81.37 \pm 0.2$	$83.16 \pm 0.2$	<b><math>86.68 \pm 0.2</math></b>	$82.30 \pm 0.2$
Fog	$35.90 \pm 0.5$	$59.76 \pm 0.3$	$59.47 \pm 0.1$	$61.08 \pm 0.3$	<b><math>61.74 \pm 0.3</math></b>
Contrast	$25.11 \pm 0.5$	$20.45 \pm 0.2$	$58.52 \pm 0.1$	$23.42 \pm 0.4$	<b><math>63.50 \pm 0.2</math></b>
Brightness	$00.98 \pm 0.6$	$70.75 \pm 0.4$	$56.32 \pm 0.3$	$16.69 \pm 0.3$	<b><math>71.84 \pm 0.2</math></b>
Saturate	$90.44 \pm 0.4$	$91.14 \pm 0.1$	$94.19 \pm 0.1$	$89.48 \pm 0.2$	<b><math>94.67 \pm 0.3</math></b>
JpegComp.	$88.21 \pm 0.2$	$89.93 \pm 0.4$	$94.64 \pm 0.1$	$85.97 \pm 0.3$	<b><math>95.89 \pm 0.4</math></b>
Elastic Trans.	$89.52 \pm 0.2$	$88.65 \pm 0.2$	$94.68 \pm 0.3$	$89.39 \pm 0.2$	<b><math>95.95 \pm 0.1</math></b>
Avg.	73.11	82.40	86.55	76.52	<b>88.20</b>
↓	16.84	10.40	08.80	16.45	<b>07.69</b>

↓ denotes that compared with the clean data (original) the average dice descent by.

TABLE II: Comparison robustness performances among different data corruption in the lung segmentation task on **JSRT** dataset using UNet as baseline. Dice is utilized as the evaluation metric.

Corruption	UNet (—)	UNet (+ VAE)	UNet (+Ours)	Rec-UNet (—)	Rec-UNet (+Ours)
Original	94.18 ± 0.2	95.43 ± 0.2	95.41 ± 0.1	95.07 ± 0.1	<b>96.56 ± 0.2</b>
Gauss. Noise	74.02 ± 0.4	75.70 ± 0.2	94.96 ± 0.1	86.26 ± 0.2	<b>95.30 ± 0.1</b>
Shot Noise	65.85 ± 0.1	74.14 ± 0.3	<b>94.53 ± 0.3</b>	78.52 ± 0.1	94.50 ± 0.3
Impulse Noise	72.78 ± 0.2	66.80 ± 0.1	<b>95.01 ± 0.1</b>	87.21 ± 0.5	94.81 ± 0.3
Speckle Noise	72.12 ± 0.4	74.34 ± 0.3	93.78 ± 0.3	82.44 ± 0.3	<b>94.67 ± 0.2</b>
Poisson Noise	76.33 ± 0.5	75.89 ± 0.1	<b>95.02 ± 0.2</b>	87.11 ± 0.1	94.83 ± 0.5
Dropout	77.50 ± 0.3	78.56 ± 0.4	94.60 ± 0.1	74.63 ± 0.2	94.83 ± 0.2
Gauss. Blur	93.38 ± 0.1	87.54 ± 0.1	95.26 ± 0.3	92.76 ± 0.4	<b>96.03 ± 0.2</b>
Glass Blur	94.13 ± 0.2	86.58 ± 0.3	94.86 ± 0.5	93.76 ± 0.2	<b>96.07 ± 0.4</b>
Defocus Blur	92.89 ± 0.3	95.44 ± 0.2	93.22 ± 0.1	86.96 ± 0.1	<b>95.87 ± 0.3</b>
Motion Blur	87.46 ± 0.2	83.77 ± 0.3	89.40 ± 0.4	87.12 ± 0.5	<b>90.06 ± 0.1</b>
Zoom Blur	<b>85.25 ± 0.5</b>	79.42 ± 0.2	82.79 ± 0.1	83.56 ± 0.3	84.66 ± 0.4
Fog	58.17 ± 0.6	64.72 ± 0.1	<b>70.45 ± 0.4</b>	62.93 ± 0.2	70.40 ± 0.2
Contrast	03.82 ± 0.2	37.59 ± 0.2	54.29 ± 0.2	03.06 ± 0.5	<b>65.80 ± 0.1</b>
Brightness	05.78 ± 0.3	42.29 ± 0.1	71.03 ± 0.3	09.54 ± 0.2	<b>74.69 ± 0.2</b>
Saturate	94.72 ± 0.5	88.15 ± 0.2	94.97 ± 0.3	93.94 ± 0.1	<b>96.18 ± 0.3</b>
JpegComp.	94.01 ± 0.1	88.14 ± 0.3	95.32 ± 0.1	93.75 ± 0.4	<b>96.23 ± 0.4</b>
Elastic Trans.	93.08 ± 0.3	93.05 ± 0.2	95.49 ± 0.3	93.45 ± 0.1	<b>95.51 ± 0.2</b>
Avgerage	67.61	75.51	88.66	76.66	<b>90.02</b>
↓	26.57	19.92	06.75	18.41	<b>06.54</b>

↓ denotes that compared with the clean data (original) the average dice descent by.

TABLE III: Comparison robustness performances among different data corruption in the lung segmentation task on **JSRT** dataset using PSPNet as baseline. Dice is utilized as the evaluation metric.

Corruption	PSPNet (—)	PSPNet (+VAE)	PSPNet (+ImageNet)	PSPNet (+Jigsaw)	PSPNet (+MoCo)	PSPNet (+Ours)	Rec-PSPNe (—)	Rec-PSPNet (+MoCo)	Rec-PSPNet (+Ours)
Original	95.34 ± 0.2	94.26 ± 0.3	96.54 ± 0.1	96.32 ± 0.3	96.50 ± 0.1	96.24 ± 0.2	95.52 ± 0.3	96.80 ± 0.1	<b>97.19 ± 0.1</b>
Gauss. Noise	84.28 ± 0.4	89.71 ± 0.2	80.97 ± 0.1	78.97 ± 0.1	87.62 ± 0.3	95.45 ± 0.2	85.35 ± 0.4	88.77 ± 0.2	<b>95.61 ± 0.1</b>
Shot Noise	88.35 ± 0.2	88.03 ± 0.1	79.51 ± 0.2	78.67 ± 0.1	87.56 ± 0.4	95.00 ± 0.2	88.53 ± 0.3	89.53 ± 0.1	<b>95.80 ± 0.2</b>
Impulse Noise	87.56 ± 0.4	89.46 ± 0.1	78.92 ± 0.4	81.40 ± 0.3	84.10 ± 0.2	95.80 ± 0.4	86.82 ± 0.3	87.08 ± 0.4	<b>95.89 ± 0.2</b>
Speckle Noise	88.64 ± 0.1	88.93 ± 0.3	79.13 ± 0.2	83.62 ± 0.1	85.63 ± 0.2	95.86 ± 0.2	88.61 ± 0.2	90.00 ± 0.2	<b>95.92 ± 0.1</b>
Poisson Noise	82.38 ± 0.3	89.11 ± 0.2	82.89 ± 0.2	78.17 ± 0.2	90.49 ± 0.3	94.51 ± 0.4	85.23 ± 0.4	91.03 ± 0.3	<b>95.01 ± 0.3</b>
Dropout	87.59 ± 0.1	86.96 ± 0.3	70.69 ± 0.2	73.19 ± 0.5	86.79 ± 0.1	94.94 ± 0.2	86.51 ± 0.1	86.06 ± 0.4	<b>95.47 ± 0.4</b>
Gauss. Blur	94.72 ± 0.2	93.23 ± 0.2	93.70 ± 0.3	93.91 ± 0.2	94.69 ± 0.3	<b>96.42 ± 0.1</b>	95.58 ± 0.1	95.54 ± 0.1	<b>96.42 ± 0.2</b>
Glass Blur	94.77 ± 0.3	93.88 ± 0.2	91.24 ± 0.1	92.83 ± 0.3	93.32 ± 0.2	<b>96.55 ± 0.2</b>	95.48 ± 0.4	95.69 ± 0.4	96.51 ± 0.1
Defocus Blur	94.87 ± 0.1	93.40 ± 0.1	93.94 ± 0.3	94.03 ± 0.3	94.95 ± 0.2	96.48 ± 0.3	95.78 ± 0.1	95.58 ± 0.3	<b>96.53 ± 0.2</b>
Motion Blur	88.97 ± 0.1	86.50 ± 0.3	88.01 ± 0.2	90.52 ± 0.4	91.76 ± 0.3	90.01 ± 0.2	89.62 ± 0.1	<b>92.74 ± 0.2</b>	90.26 ± 0.1
Zoom Blur	82.65 ± 0.4	83.87 ± 0.2	85.88 ± 0.3	84.91 ± 0.1	89.11 ± 0.2	84.47 ± 0.1	85.44 ± 0.2	<b>89.34 ± 0.1</b>	85.23 ± 0.2
Fog	58.48 ± 0.2	61.80 ± 0.1	77.58 ± 0.4	76.30 ± 0.2	<b>79.77 ± 0.2</b>	74.50 ± 0.3	61.68 ± 0.3	79.55 ± 0.2	76.48 ± 0.1
Contrast	43.77 ± 0.5	60.73 ± 0.3	59.07 ± 0.2	56.59 ± 0.4	65.55 ± 0.4	65.37 ± 0.3	55.71 ± 0.4	65.47 ± 0.2	<b>67.74 ± 0.3</b>
Brightness	00.75 ± 0.1	72.73 ± 0.2	81.99 ± 0.2	82.08 ± 0.3	82.93 ± 0.3	79.89 ± 0.1	01.27 ± 0.1	<b>83.78 ± 0.1</b>	81.95 ± 0.1
Saturate	95.34 ± 0.2	94.26 ± 0.1	96.52 ± 0.1	96.32 ± 0.1	96.47 ± 0.1	96.18 ± 0.1	96.32 ± 0.2	96.42 ± 0.3	<b>97.11 ± 0.2</b>
JpegComp.	95.16 ± 0.1	94.09 ± 0.1	91.22 ± 0.4	91.47 ± 0.3	92.88 ± 0.3	96.30 ± 0.3	96.11 ± 0.1	93.50 ± 0.3	<b>96.88 ± 0.3</b>
Elastic Trans.	95.15 ± 0.4	94.08 ± 0.1	94.07 ± 0.3	95.32 ± 0.1	96.65 ± 0.2	96.55 ± 0.1	95.90 ± 0.3	96.27 ± 0.2	<b>96.74 ± 0.1</b>
Avgerage	80.20	85.93	83.84	84.01	88.25	90.65	81.76	89.20	<b>91.51</b>
↓	15.14	08.33	12.80	12.31	07.99	05.59	13.76	07.60	<b>05.68</b>

↓ denotes that compared with the clean data (original) the average dice descent by.

TABLE IV: Comparison robustness performances among different data corruption in the lung segmentation task on **SH** dataset using UNet as baseline. Dice is utilized as the evaluation metric.

Corruption	UNet (—)	UNet (+ VAE)	UNet (+Ours)	Rec-UNet (—)	Rec-UNet (+Ours)
Original	88.03 ± 0.2	86.66 ± 0.2	91.31 ± 0.2	89.34 ± 0.4	<b>93.17 ± 0.3</b>
Gauss. Noise	85.97 ± 0.4	86.47 ± 0.2	<b>90.50 ± 0.2</b>	86.59 ± 0.2	90.37 ± 0.1
Shot Noise	83.78 ± 0.3	84.93 ± 0.4	<b>90.61 ± 0.3</b>	83.09 ± 0.2	88.45 ± 0.2
Impulse Noise	86.96 ± 0.2	85.57 ± 0.1	90.18 ± 0.4	87.09 ± 0.3	<b>90.22 ± 0.3</b>
Speckle Noise	82.03 ± 0.2	82.91 ± 0.4	87.64 ± 0.2	81.31 ± 0.1	<b>88.81 ± 0.3</b>
Poisson Noise	84.03 ± 0.3	84.23 ± 0.4	87.73 ± 0.3	83.60 ± 0.1	<b>88.36 ± 0.1</b>
Dropout	88.30 ± 0.3	86.17 ± 0.2	<b>89.03 ± 0.3</b>	84.24 ± 0.2	88.99 ± 0.2
Gauss. Blur	85.07 ± 0.1	84.86 ± 0.2	89.47 ± 0.3	85.79 ± 0.2	<b>89.93 ± 0.2</b>
Glass Blur	85.25 ± 0.2	86.43 ± 0.4	90.76 ± 0.1	85.79 ± 0.1	<b>91.56 ± 0.2</b>
Defocus Blur	85.05 ± 0.1	86.66 ± 0.4	<b>90.90 ± 0.2</b>	86.53 ± 0.2	90.05 ± 0.4
Motion Blur	80.73 ± 0.2	81.62 ± 0.2	84.91 ± 0.5	82.45 ± 0.2	<b>85.56 ± 0.3</b>
Zoom Blur	69.43 ± 0.3	71.28 ± 0.1	77.68 ± 0.2	70.19 ± 0.1	<b>79.06 ± 0.4</b>
Fog	55.31 ± 0.4	51.22 ± 0.4	56.70 ± 0.2	53.25 ± 0.2	<b>56.98 ± 0.4</b>
Contrast	37.75 ± 0.2	35.45 ± 0.2	<b>43.19 ± 0.2</b>	35.85 ± 0.3	41.78 ± 0.1
Brightness	57.72 ± 0.4	61.48 ± 0.1	62.70 ± 0.2	<b>65.68 ± 0.3</b>	65.41 ± 0.1
Saturate	87.64 ± 0.4	88.93 ± 0.4	<b>91.72 ± 0.3</b>	86.90 ± 0.2	90.96 ± 0.1
JpegComp.	87.55 ± 0.4	88.44 ± 0.2	91.34 ± 0.2	87.72 ± 0.1	<b>91.76 ± 0.2</b>
Elastic Trans.	86.07 ± 0.2	86.47 ± 0.4	91.48 ± 0.2	86.82 ± 0.2	<b>92.95 ± 0.3</b>
Avgerage	78.16	78.42	<b>84.50</b>	78.41	83.01
↓	09.87	08.24	<b>06.81</b>	10.93	10.17

↓ denotes that compared with the clean data (original) the average dice descent by.

TABLE V: Comparison robustness performances among different data corruption in the lung segmentation task on **SH** dataset using PSPNet as baseline. Dice is utilized as the evaluation metric.

Corruption	PSPNet (—)	PSPNet (+VAE)	PSPNet (+ImageNet)	PSPNet (+Jigsaw)	PSPNet (+MoCo)	PSPNet (+Ours)	Rec-PSPNet (—)	Rec-PSPNet (+MoCo)	Rec-PSPNet (+Ours)
Original	83.26 ± 0.3	91.97 ± 0.2	92.67 ± 0.2	93.95 ± 0.1	94.74 ± 0.1	92.68 ± 0.3	85.17 ± 0.3	<b>94.80 ± 0.2</b>	94.77 ± 0.2
Gauss. Noise	82.45 ± 0.2	83.43 ± 0.1	85.06 ± 0.2	86.99 ± 0.4	87.46 ± 0.2	85.93 ± 0.4	84.32 ± 0.1	90.38 ± 0.2	<b>92.03 ± 0.2</b>
Shot Noise	82.72 ± 0.1	76.10 ± 0.3	85.57 ± 0.3	85.38 ± 0.2	87.10 ± 0.4	84.96 ± 0.1	83.02 ± 0.3	89.95 ± 0.2	<b>91.00 ± 0.2</b>
Impulse Noise	81.46 ± 0.2	85.22 ± 0.3	85.61 ± 0.1	88.67 ± 0.4	85.48 ± 0.3	86.28 ± 0.2	83.09 ± 0.3	88.36 ± 0.1	<b>92.16 ± 0.3</b>
Speckle Noise	82.77 ± 0.2	83.72 ± 0.3	83.02 ± 0.4	83.08 ± 0.1	86.17 ± 0.3	84.20 ± 0.2	81.96 ± 0.3	86.33 ± 0.1	<b>90.00 ± 0.1</b>
Poisson Noise	79.89 ± 0.2	53.77 ± 0.4	88.43 ± 0.3	78.31 ± 0.2	89.63 ± 0.1	84.44 ± 0.3	83.88 ± 0.1	<b>91.65 ± 0.3</b>	90.88 ± 0.2
Dropout	79.48 ± 0.2	67.69 ± 0.2	77.57 ± 0.1	81.24 ± 0.3	76.35 ± 0.1	82.73 ± 0.2	75.57 ± 0.3	80.34 ± 0.2	<b>86.96 ± 0.1</b>
Gauss. Blur	84.77 ± 0.3	90.65 ± 0.2	87.08 ± 0.4	78.56 ± 0.1	91.01 ± 0.3	84.96 ± 0.2	85.34 ± 0.1	88.46 ± 0.2	<b>91.03 ± 0.1</b>
Glass Blur	85.32 ± 0.3	90.92 ± 0.3	83.21 ± 0.1	88.44 ± 0.2	91.19 ± 0.3	85.56 ± 0.1	85.44 ± 0.3	91.00 ± 0.1	<b>91.35 ± 0.2</b>
Defocus Blur	83.54 ± 0.2	90.80 ± 0.2	87.23 ± 0.2	80.63 ± 0.4	91.68 ± 0.1	85.45 ± 0.3	85.48 ± 0.2	<b>93.52 ± 0.2</b>	91.29 ± 0.3
Motion Blur	81.56 ± 0.1	85.72 ± 0.3	82.58 ± 0.2	83.63 ± 0.4	<b>88.81 ± 0.3</b>	85.72 ± 0.2	81.75 ± 0.2	88.51 ± 0.2	86.49 ± 0.1
Zoom Blur	76.80 ± 0.2	76.82 ± 0.2	81.90 ± 0.3	80.25 ± 0.1	87.90 ± 0.4	81.52 ± 0.3	72.28 ± 0.2	<b>89.02 ± 0.3</b>	83.84 ± 0.2
Fog	57.87 ± 0.2	54.68 ± 0.3	60.73 ± 0.1	55.27 ± 0.2	74.83 ± 0.3	60.79 ± 0.4	54.55 ± 0.3	76.28 ± 0.2	<b>86.79 ± 0.2</b>
Contrast	40.27 ± 0.3	56.61 ± 0.3	51.69 ± 0.4	43.87 ± 0.1	54.16 ± 0.3	53.06 ± 0.4	59.10 ± 0.1	58.72 ± 0.3	<b>60.04 ± 0.2</b>
Brightness	00.11 ± 0.3	65.60 ± 0.3	64.86 ± 0.2	<b>67.40 ± 0.4</b>	63.60 ± 0.4	65.68 ± 0.1	06.60 ± 0.2	64.09 ± 0.1	64.54 ± 0.3
Saturate	83.22 ± 0.2	91.91 ± 0.1	93.96 ± 0.4	93.93 ± 0.3	<b>95.75 ± 0.2</b>	90.44 ± 0.3	84.10 ± 0.3	94.10 ± 0.2	93.00 ± 0.1
JpegComp.	82.95 ± 0.1	83.43 ± 0.1	91.73 ± 0.1	87.85 ± 0.3	85.98 ± 0.2	89.13 ± 0.4	85.85 ± 0.1	89.63 ± 0.3	<b>92.81 ± 0.2</b>
Elastic Trans.	83.32 ± 0.1	91.19 ± 0.2	89.20 ± 0.4	91.33 ± 0.4	<b>94.38 ± 0.3</b>	86.97 ± 0.5	84.98 ± 0.3	93.58 ± 0.2	92.46 ± 0.1
Avgerage	68.77	78.13	81.14	79.68	84.20	81.05	75.14	85.52	<b>86.86</b>
↓	14.49	13.84	11.53	14.30	10.54	11.63	10.03	09.28	<b>07.91</b>

↓ denotes that compared with the clean data (original) the average dice descent by.

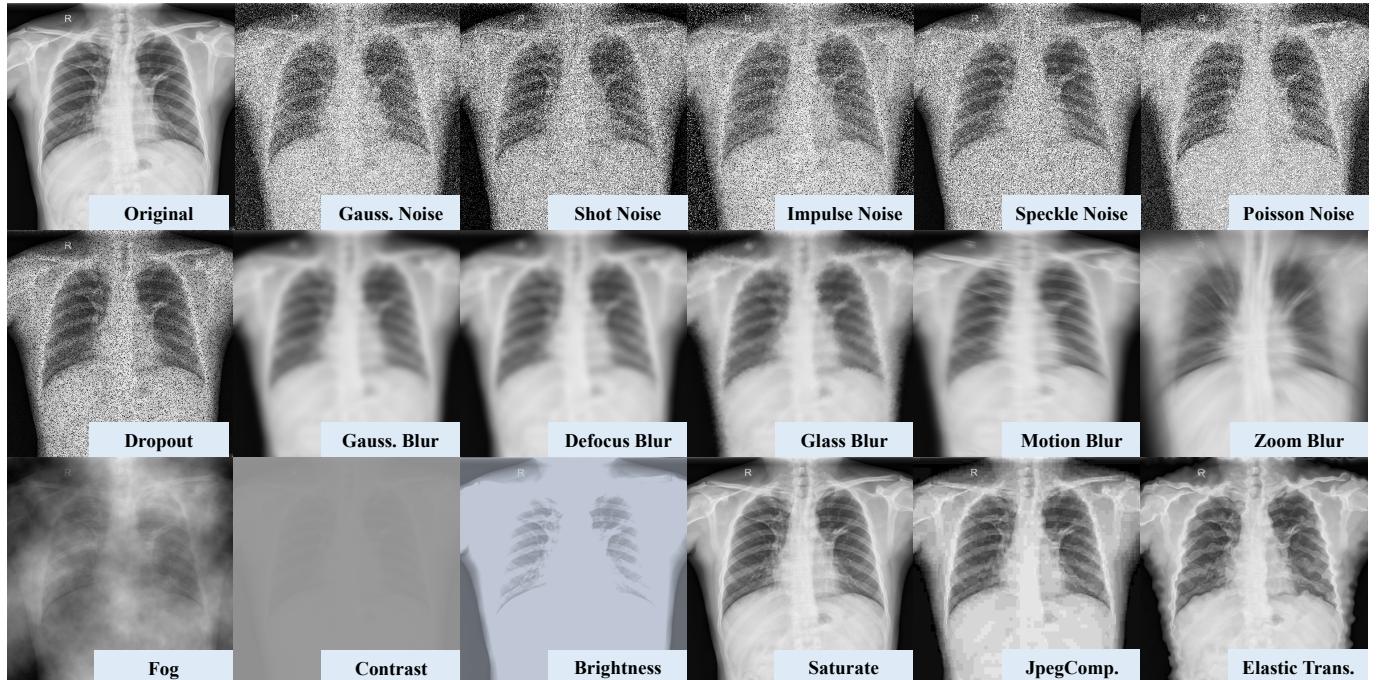


Fig. 1: The visualization of different data corruptions and perturbations. We choose these noises, blur, and digital transformations that are common in real-world applications for verification of the robustness of our model.

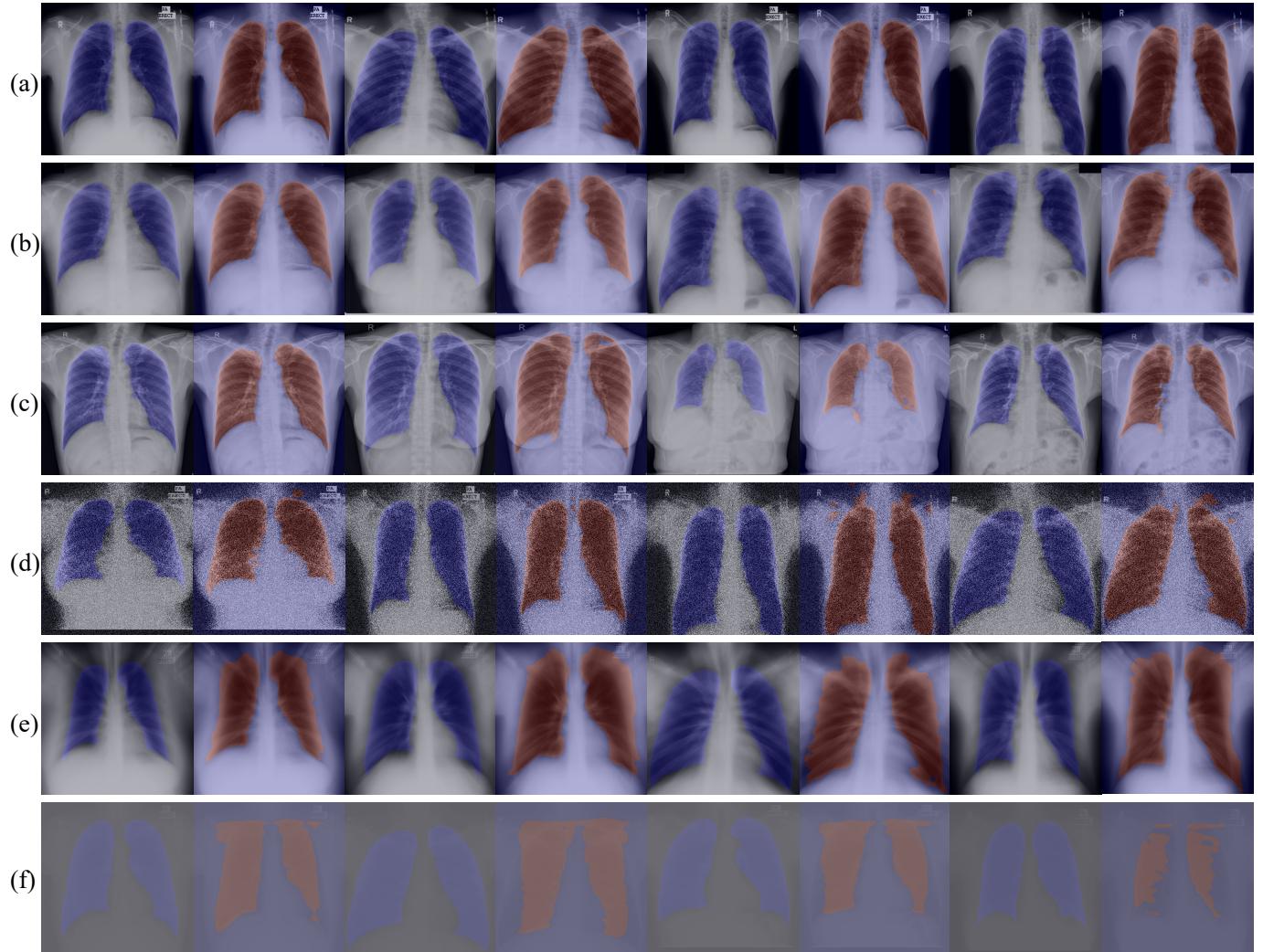


Fig. 2: The visualization of segmentation results. The blue region denotes the mask, and the red area is the predicted result. The model is trained on M dataset, then tests on (a) MC validation dataset; (b) JSRT dataset; (c) SH dataset; (d) Poisson noised MC dataset; (e) zoom blurred MC dataset; (f) contrast transformed MC dataset.