Attack Generalization of the Generator

We demonstrate that our generator G_θ can generalize the attack across the dataset in Table 1 and 2. Table 1 shows the performance of the attack if generator is trained on SHHA and tested on UCF-QNRF, and table 2 shows the case trained pn UCF-QNRF and tested on SHHA.

Table 1: Transferability of our designed adversarial examples across crowd counting models on UCF-QNRF dataset, where G_θ was trained on SHHA

Source Model	**Target Model**						
	SASNet	**P2PNet**	**FIDTM**	**PET**	**HMoDE**	**APGCC**	**Gramformer**
SASNet	620.13	819.79	741.86	756.81	572.13	808.23	613.94
P2PNet	606.15	657.84	622.34	676.95	429.78	673.17	518.40
FIDTM	639.64	833.09	767.54	766.21	530.23	819.93	647.01
PET	630.66	830.98	751.50	761.65	526.67	808.33	622.28
HMoDE	619.76	820.67	741.68	757.75	528.11	807.53	613.05
APGCC	659.61	840.08	765.43	786.79	561.54	816.36	637.51
Gramformer	617.21	819.49	736.08	761.39	524.83	806.27	609.38
Clean	326.94	324.73	279.08	393.19	339.02	288.12	290.44

Table 2: Transferability of our designed adversarial examples across crowd counting models on SHHA dataset, where G_θ was trained on UCF-QNRF

Source Model	**Target Model**						
	SASNet	**P2PNet**	**FIDTM**	**PET**	**HMoDE**	**APGCC**	**Gramformer**
SASNet	300.72	420.65	338.49	293.55	250.26	399.21	257.62
P2PNet	262.40	296.66	259.92	230.87	171.15	308.70	200.62
FIDTM	300.92	421.37	338.45	293.33	249.61	398.67	257.68
PET	305.91	426.73	345.73	299.04	251.48	405.39	262.66
HMoDE	300.89	420.34	338.98	293.14	249.70	398.38	257.52
APGCC	308.81	422.38	339.53	294.61	251.20	353.59	259.73
Gramformer	300.64	420.63	339.05	292.36	250.11	398.36	257.63
Clean	36.98	46.12	28.10	70.21	74.80	56.55	29.91

Benchmark Performance on UCF-QNRF

We additionally compute the performance of our proposed method with state-of-the-art benchmark methods in Table 3. We observe that we have achieved the balance between higher overall PSNR and SSIM with stronger attack higher MAE and MR.

Table 3: Evaluation of attacks using the FIDTM model across different density regimes on UCF-QNRF. Higher is better (↑).

(a) Overall

Method	MAE↑	MR↑	SSIM↑	PSNR↑
Clean	279.08	18.73	N/A	N/A
DI^{2}FGSM	532.50	33.01	0.77	18.73
Admix	520.34	32.03	0.78	18.75
FIA	520.28	31.87	0.77	18.72
PAP	296.00	20.17	0.95	24.28
SVRE	530.57	32.74	0.77	18.72
GRA	518.30	31.63	0.78	18.74
L2T	521.98	32.19	0.78	18.77
GE-AdvGAN	502.39	32.27	0.84	22.86
DiffAttack	893.02	71.66	0.49	11.34
Ours (P2PNet)	632.00	51.72	0.82	19.35
Ours (SASNet)	741.65	61.39	0.79	17.93

(b) Sparse (<100)

Method	MAE↑	MR↑	SSIM↑	PSNR↑
Clean	16.50	20.34	N/A	N/A
DI^{2}FGSM	11.00	13.74	0.81	20.00
Admix	12.33	15.46	0.81	20.05
FIA	12.00	14.27	0.81	19.99
PAP	16.50	20.53	0.95	26.30
SVRE	10.50	13.23	0.81	19.99
GRA	10.67	13.12	0.81	20.02
L2T	11.00	13.64	0.81	20.05
GE-AdvGAN	8.17	10.74	0.86	23.28
DiffAttack	30.00	36.16	0.55	11.89
Ours (P2PNet)	18.83	23.93	0.78	18.64
Ours (SASNet)	24.33	30.42	0.75	17.10

(c) Moderate (100–1000)

Method	MAE↑	MR↑	SSIM↑	PSNR↑
Clean	73.32	15.69	N/A	N/A
DI^{2}FGSM	128.05	25.50	0.80	19.32
Admix	122.70	24.61	0.80	19.35
FIA	122.25	24.46	0.80	19.32
PAP	81.07	17.36	0.94	24.15
SVRE	127.72	25.24	0.80	19.32
GRA	121.43	24.21	0.80	19.34
L2T	124.05	24.87	0.80	19.36
GE-AdvGAN	130.76	26.59	0.85	23.21
DiffAttack	308.53	64.74	0.53	11.70
Ours (P2PNet)	229.62	48.47	0.82	19.57
Ours (SASNet)	273.96	57.52	0.80	18.14

(d) Dense (>1000)

Method	MAE↑	MR↑	SSIM↑	PSNR↑
Clean	773.03	25.66	N/A	N/A
DI^{2}FGSM	1503.80	51.68	0.72	17.26
Admix	1474.99	50.34	0.72	17.28
FIA	1475.84	50.21	0.72	17.26
PAP	812.33	26.65	0.95	24.44
SVRE	1498.07	51.39	0.72	17.26
GRA	1471.11	50.06	0.72	17.27
L2T	1477.46	50.38	0.72	17.29
GE-AdvGAN	1395.88	46.87	0.82	22.01
DiffAttack	2320.09	90.29	0.41	10.45
Ours (P2PNet)	1604.69	61.11	0.81	18.90
Ours (SASNet)	1872.53	72.45	0.78	17.50

Adversarial Examples

Below are some of the adversarial examples and perturbations shown in Figure 1 that are generated using our proposed method.

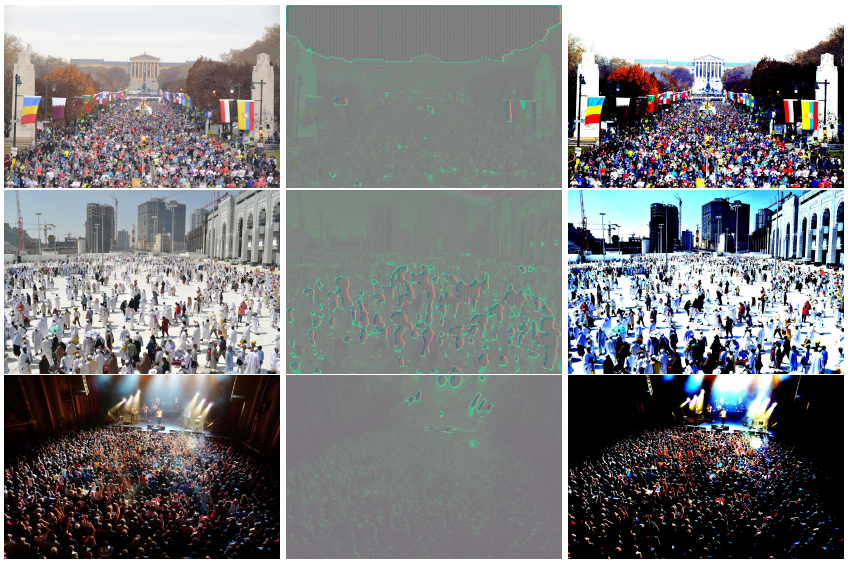

Figure 1: Left: Clean Image, Middle: Perturbation, Right: Adversarial Image

Benchmark Examples Comparison

Below we compare adversarial examples generated from our image with DiffAttack and PAP for qualitative comparison in Figure 2.

Implementation Codes

The code for implementing the attack described in the paper is included, full code will be released upon acceptance and publication.

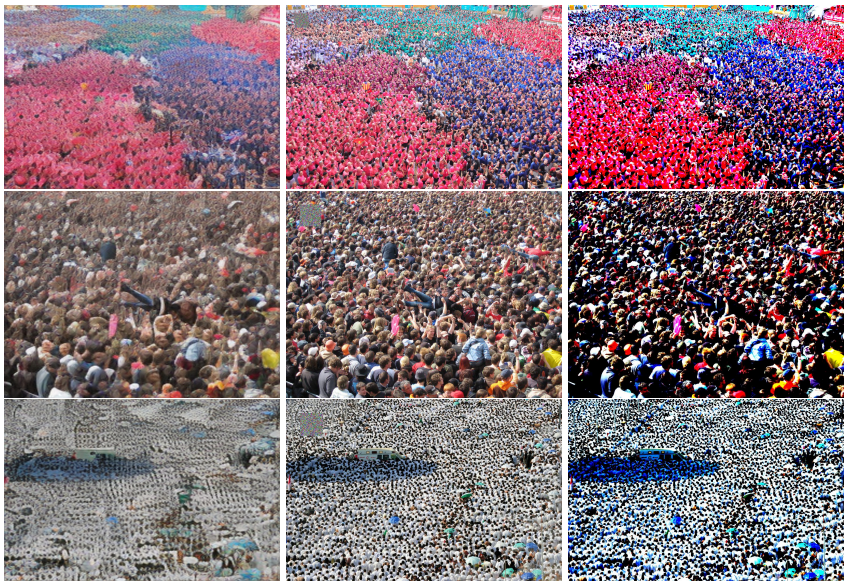

Figure 2: Left: DiffAttack, Middle: PAP, Right: Ours