Supplementary File for:
“Painting from Part”

Dongsheng Guo[1] Haoru Zhao[1] Yunhao Cheng[1] Haiyong Zheng[1,*] Zhaorui Gu[1] Bing Zheng[1,2]
[1]Underwater Vision Lab (http://ouc.ai), Ocean University of China
[2]Sanya Oceanographic Institution, Ocean University of China

A. Implementation Details

Our part encoder consists of six stride-2 partial convolutions [9] and two sibling fully-connected (FC) layers to extract part-features and produce part-noise. Our painting generator is composed of five repainting residual blocks and a normal residual blocks. We adopt the similar structure to DenseNet [4] as our whole discriminator and multi-scale PatchGAN [6] as our painting discriminator. Our network is trained using Adam optimizer [8] with parameters of $\beta_1 = 0.0$, $\beta_2 = 0.9$, learning rate $\alpha_1 = 0.0001$ for part encoder and painting generator, and learning rate $\alpha_2 = 0.0004$ for whole and painting discriminators.

Figure A shows the detailed architecture of our Repainting Residual Block (RRB). Table A, Table B, Table C, and Table D present the detailed architectures of our part encoder, painting generator, whole discriminator, and painting discriminator, respectively.

Layer ID	Type	Norm	Act.	K	S	P	Out
1	PConv	IN	LReLU	3	2	1	64
2	PConv	IN	LReLU	3	2	1	128
3	PConv	IN	LReLU	3	2	1	256
4	PConv	IN	LReLU	3	2	1	512
5	PConv	IN	LReLU	3	2	1	512
6	PConv	IN	LReLU	3	2	1	512
7	FC | FC	–	–	–	–	–	256

Table A. **The architecture of our part encoder.** “Norm” denotes normalization type after convolution, “Act.” denotes activation type, “K” denotes kernel size, “S” denotes stride, “P” denotes padding, and “Out” denotes output channel number. “PConv” means partial convolution [9], “IN” means instance normalization [13], and “LReLU” means Leaky ReLU [10] with a negative slope of 0.2.

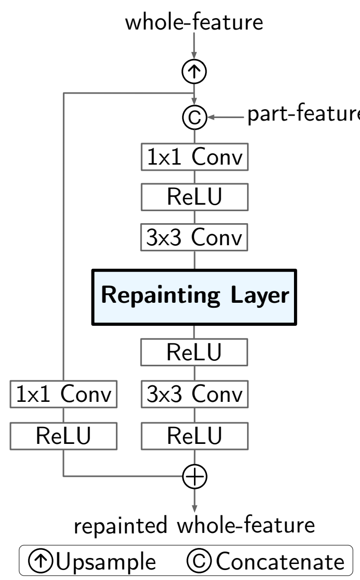


Figure A. The detailed architecture of our Repainting Residual Block (RRB).

Layer ID	Type	Norm	Act.	K	S	P	Out
1	FC	–	–	–	–	–	16384
2	RRB	–	–	–	–	–	1024
3	RRB	–	–	–	–	–	1024
4	RRB	–	–	–	–	–	512
5	RRB	–	–	–	–	–	256
6	RRB	–	–	–	–	–	128
7	RB	–	–	–	–	–	64
8	PRM	–	–	–	–	–	64
9	Conv	–	Tanh	3	1	1	3

Table B. **The architecture of our painting generator and the following structure.** “Norm” denotes normalization type after convolution, “Act.” denotes activation type, “K” denotes kernel size, “S” denotes stride, “P” denotes padding, and “Out” denotes output channel number. “RRB” means repainting residual block, “RB” means normal residual block (“RRB” without repainting layer), and “PRM” means part-patch refining module. Please refer to Figures 3 and 4 in the paper for details of RRB and PRM respectively.

*Corresponding author: Haiyong Zheng (zhenghaiyong@ouc.edu.cn).
This work was supported by the National Natural Science Foundation of China under Grant Nos. 61771440 and 41776113, and the Fundamental Research Funds for the Central Universities under Grant No. 202061002.

Layer ID	Type	Skip	Norm	Act.	K	S	P	Out
1	Conv	–	BN	ReLU	7	2	3	64
2	Max Pool	–	BN	ReLU	3	2	1	64
3	Conv	–	BN	ReLU	1	1	0	128
4	Conv	2	BN	ReLU	3	1	1	96
5	Conv	–	BN	ReLU	1	1	0	128
6	Conv	2,4	BN	ReLU	3	1	1	128
7	Conv	–	BN	ReLU	1	1	0	128
8	Conv	2,4,6	BN	ReLU	3	1	1	160
9	Conv	–	–	–	1	1	0	80
10	Average Pool	–	BN	ReLU	2	2	0	80
11	Conv	–	BN	ReLU	1	1	0	128
12	Conv	10	BN	ReLU	3	1	1	112
13	Conv	–	BN	ReLU	1	1	0	128
14	Conv	10,12	BN	ReLU	3	1	1	144
15	Conv	–	BN	ReLU	1	1	0	128
16	Conv	10,12,14	BN	ReLU	3	1	1	176
17	Conv	–	–	–	1	1	0	88
18	Average Pool	–	BN	ReLU	2	2	0	88
19	Conv	–	BN	ReLU	1	1	0	128
20	Conv	18	BN	ReLU	3	1	1	120
21	Conv	–	BN	ReLU	1	1	0	128
22	Conv	18,20	BN	ReLU	3	1	1	152
23	Conv	–	BN	ReLU	1	1	0	128
24	Conv	18,20,22	BN	ReLU	3	1	1	184
25	Average Pool	–	–	Sigmoid	7	1	0	184

Table C. **The architecture of our whole discriminator.** “Skip” denotes the layer IDs of concatenated feature maps before normalization, “Norm” denotes normalization type, “Act.” denotes activation type, “K” denotes kernel size, “S” denotes stride, “P” denotes padding, and “Out” denotes output channel number. “BN” means batch normalization [5].

Layer ID	Type	Norm	Act.	K	S	P	Out
Scale I							
1	Conv	SN	LReLU	4	2	2	64
2	Conv	SN	LReLU	4	2	2	128
3	Conv	SN	LReLU	4	2	2	256
4	Conv	SN	LReLU	4	1	2	512
5	Conv	SN	Sigmoid	4	1	2	1
Scale II							
1	Average Pool	–	–	3	2	1	–
2	Conv	SN	LReLU	4	2	2	64
3	Conv	SN	LReLU	4	2	2	128
4	Conv	SN	LReLU	4	2	2	256
5	Conv	SN	LReLU	4	1	2	512
6	Conv	SN	Sigmoid	4	1	2	1

Table D. **The architecture of our painting discriminator.** “Norm” denotes normalization type after convolution, “Act.” denotes activation type, “K” denotes kernel size, “S” denotes stride, “P” denotes padding, and “Out” denotes output channel number. “SN” means spectral normalization [11], and “LReLU” means Leaky ReLU [10] with a negative slope of 0.2.

B. Dataset Splitting

We list dataset splitting for training and testing in Table E, where, we keep default official split on AFHQ Cat, Cityscapes, and Places2, while we randomly select samples on CelebA-HQ, CUB, Flowers, and Paris StreetView.

Dataset	Training	Testing	Total
Inpainting			
CelebA-HQ [7]	$28,000$	$2,000$	$30,000$
Places2 (10 Categories) [15]	$50,000$	100	$50,100$
Regular Outpainting			
CelebA-HQ [7]	$28,000$	$2,000$	$30,000$
CUB [14]	$4,200$	915	$5,115$
AFHQ Cat [1]	$5,153$	500	$5,653$
Flowers [12]	$7,000$	$1,189$	$8,189$
Paris StreetView [3]	$13,000$	$1,900$	$14,900$
Cityscapes [2]	$2,975$	$1,525$	$4,500$
Places2 Desert Road [15]	$5,000$	100	$5,100$
Irregular Outpainting			
CelebA-HQ [7]	$28,000$	$2,000$	$30,000$
Places2 (10 Categories) [15]	$50,000$	100	$50,100$

Table E. **Dataset splitting for training and testing in inpainting, regular outpainting, irregular outpainting.** For Places2 of inpainting and irregular outpainting, we choose the following 10 categories for training: “valley”, “sky”, “desert road”, “highway”, “mountain”, “mountain snow”, “mountain path”, “snowfield”, “lake natural”, and “river”, while we use “valley” category for testing.

C. Additional Qualitative Results

We show additional qualitative results of inpainting in Figures B and C, regular outpainting in Figures D–J, and irregular outpainting on different datasets in Figures K and L.

References

[1] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. StarGAN v2: Diverse image synthesis for multiple domains. In *CVPR*, pages 8188–8197, 2020. 2, 8

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, pages 3213–3223, 2016. 2, 11

[3] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A Efros. What makes paris look like paris? *ACM TOG*, 31(4):101, 2012. 2, 10

[4] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, pages 4700–4708, 2017. 1

[5] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. 2

[6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, pages 1125–1134, 2017. 1

[7] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *ICLR*, 2018. 2, 4, 6, 13

[8] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 1

[9] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for ir-

regular holes using partial convolutions. In *ECCV*, pages 85–100, 2018. 1

[10] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *ICMLW*, pages 1–6, 2013. 1, 2

[11] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *ICLR*, 2018. 2

[12] M-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In *ICVGIP*, pages 722–729, 2008. 2, 9

[13] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 1

[14] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 2, 7

[15] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE TPAMI*, pages 1452–1464, 2017. 2, 5, 12, 14

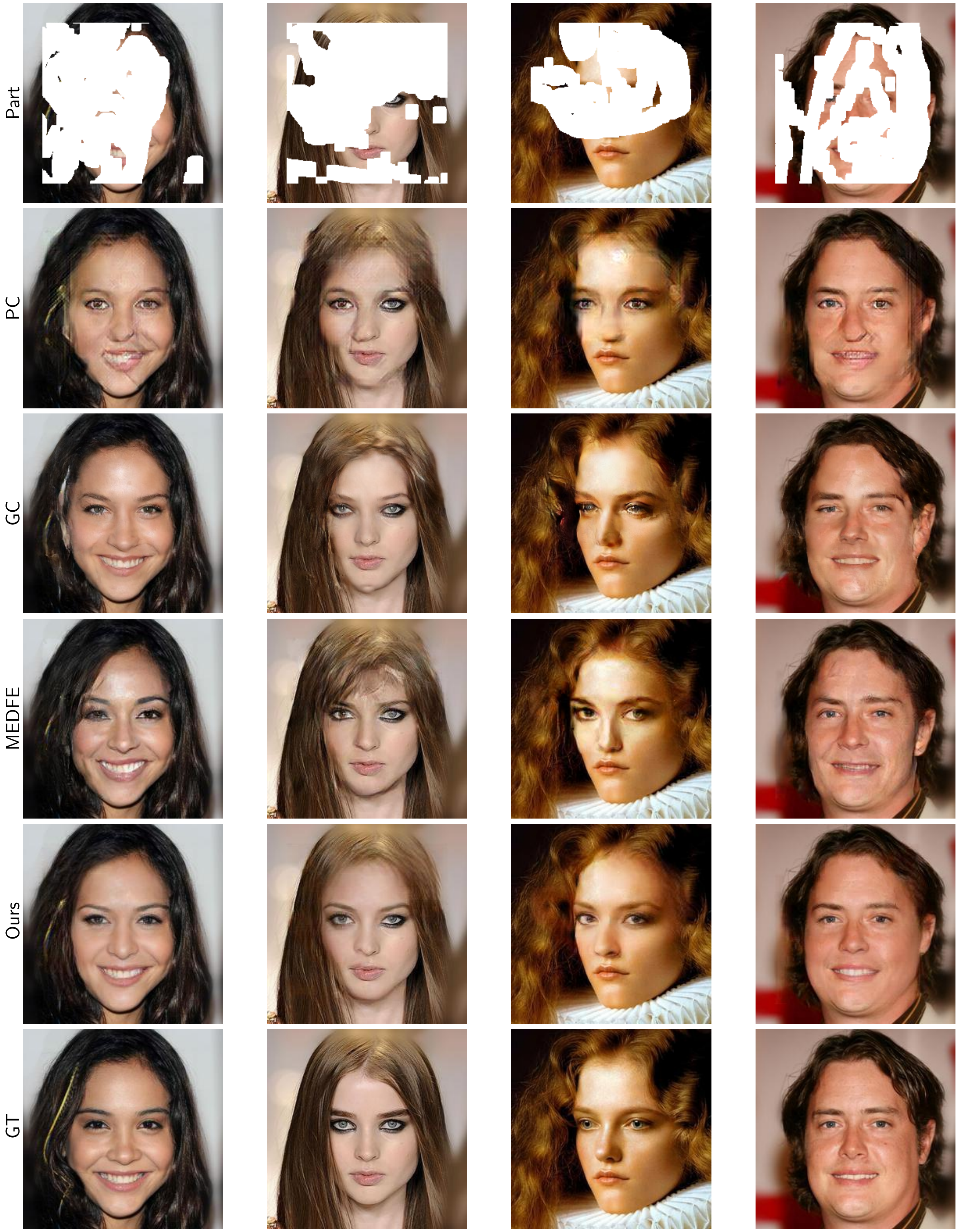


Figure B. Additional qualitative results of image inpainting on CelebA-HQ [7].

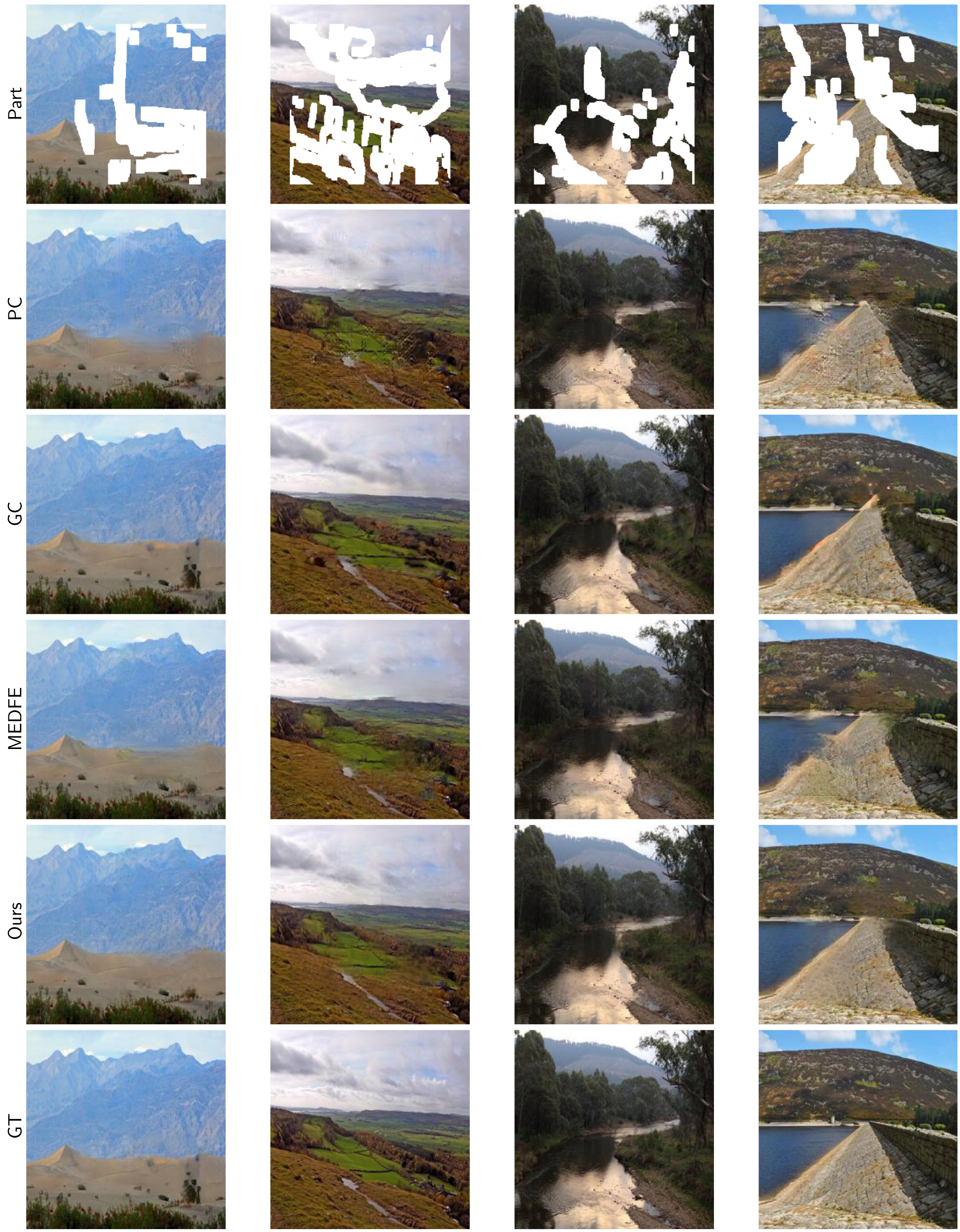
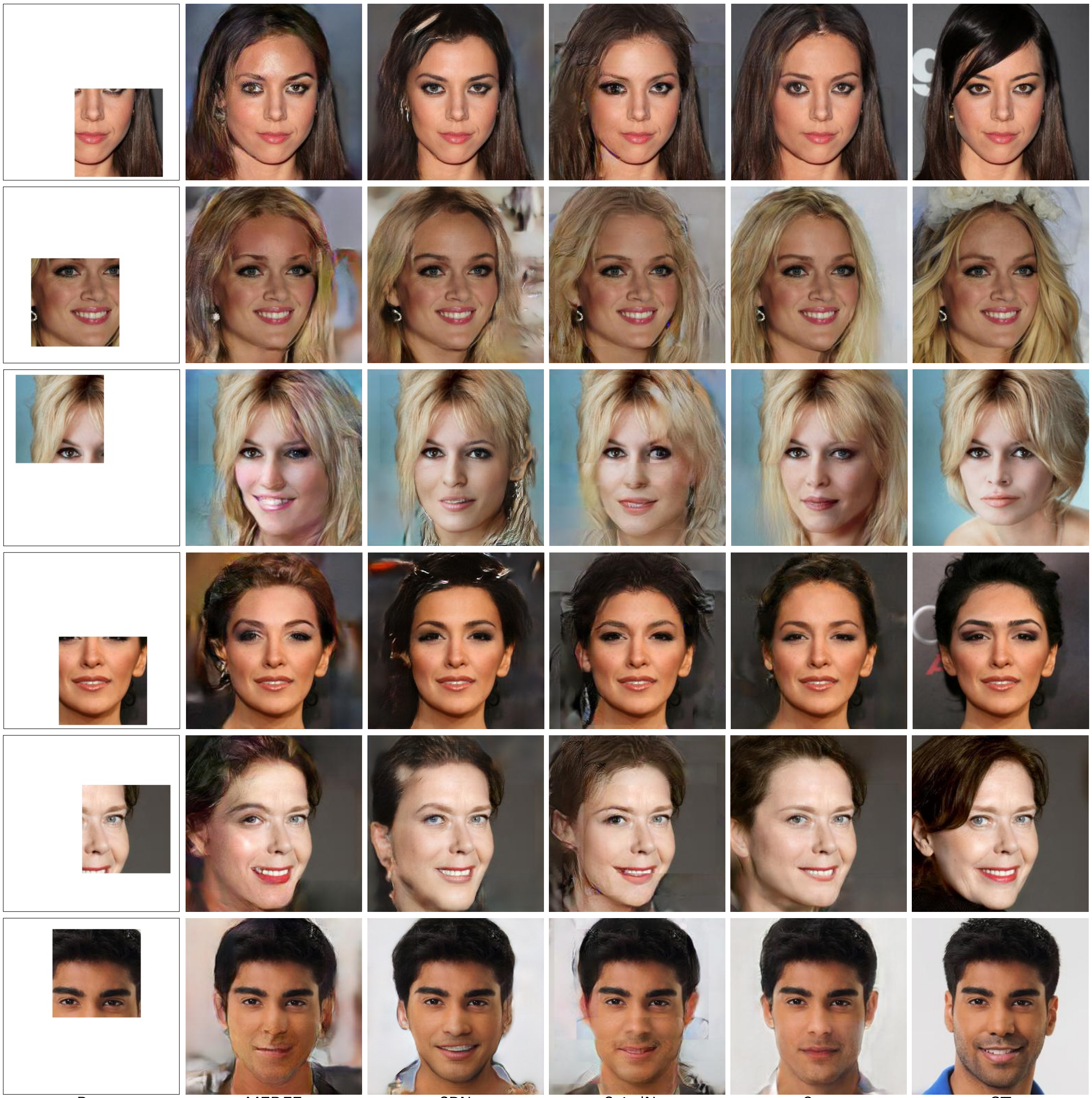


Figure C. Additional qualitative results of image inpainting on Places2 [15].



Part MEDFE SRN SpiralNet Ours GT

Figure D. Additional qualitative results of regular image outpainting on CelebA-HQ [7].

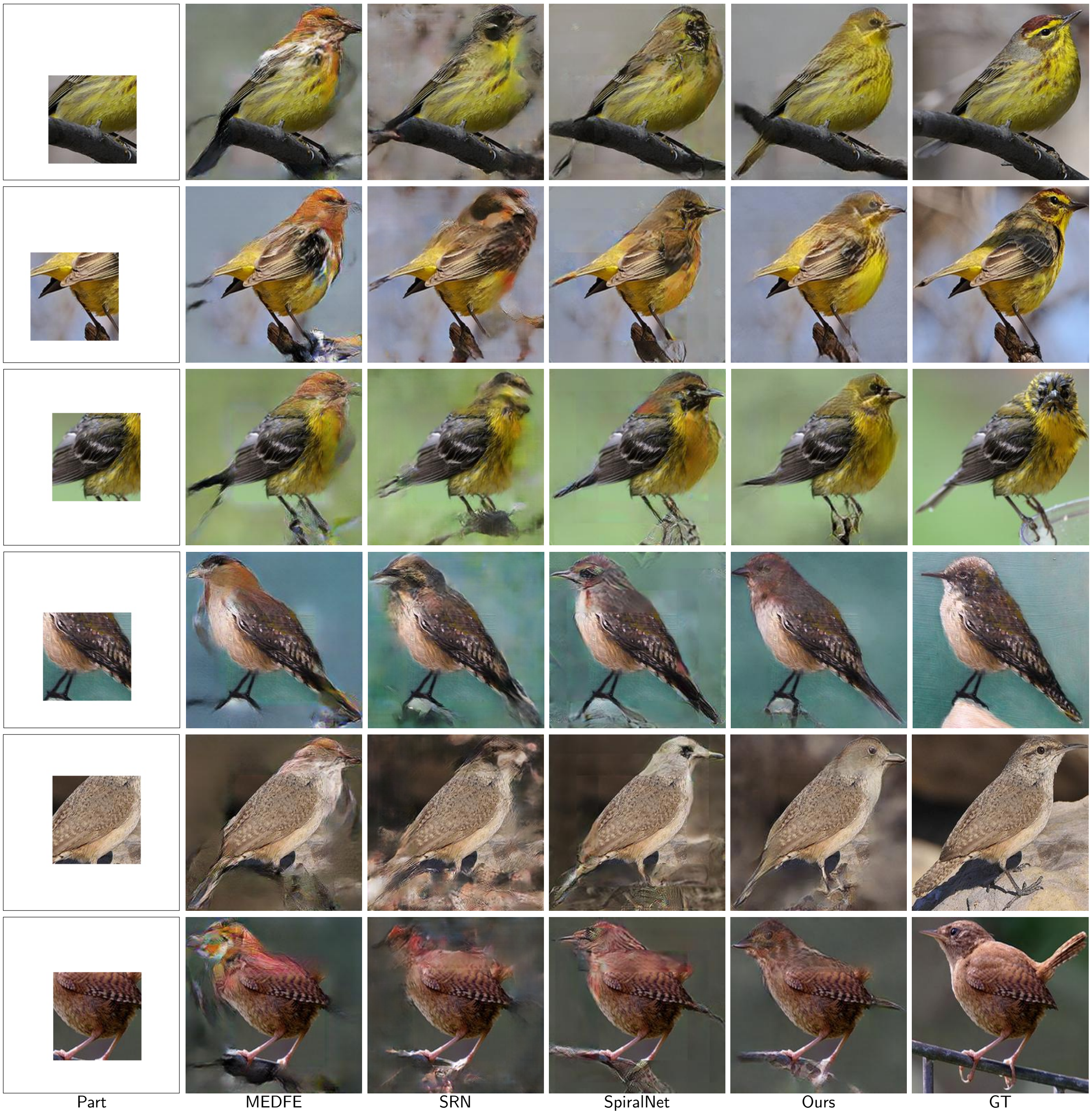


Figure E. Additional qualitative results of regular image outpainting on CUB [14].

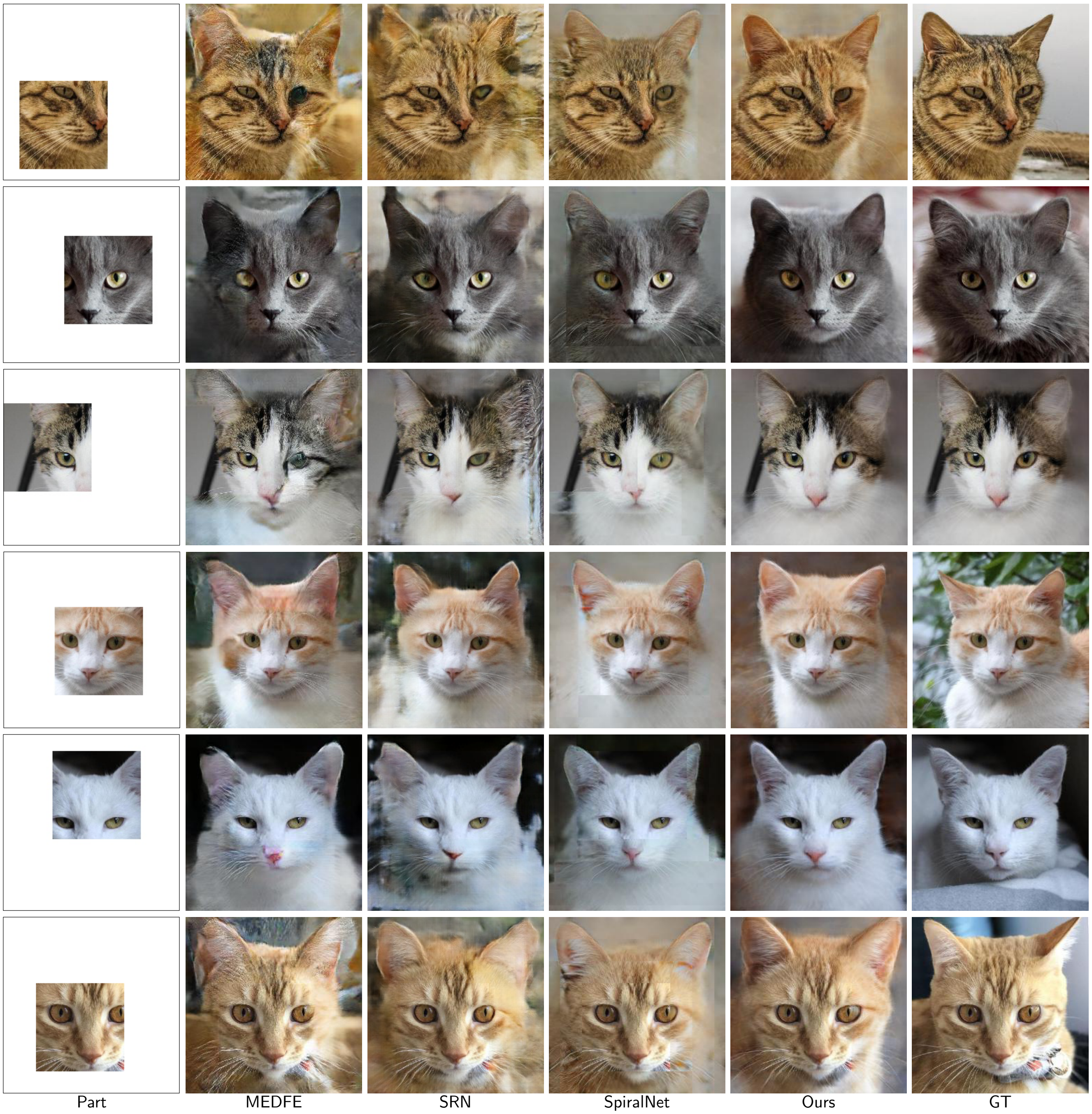


Figure F. Additional qualitative results of regular image outpainting on AFHQ Cat [1].

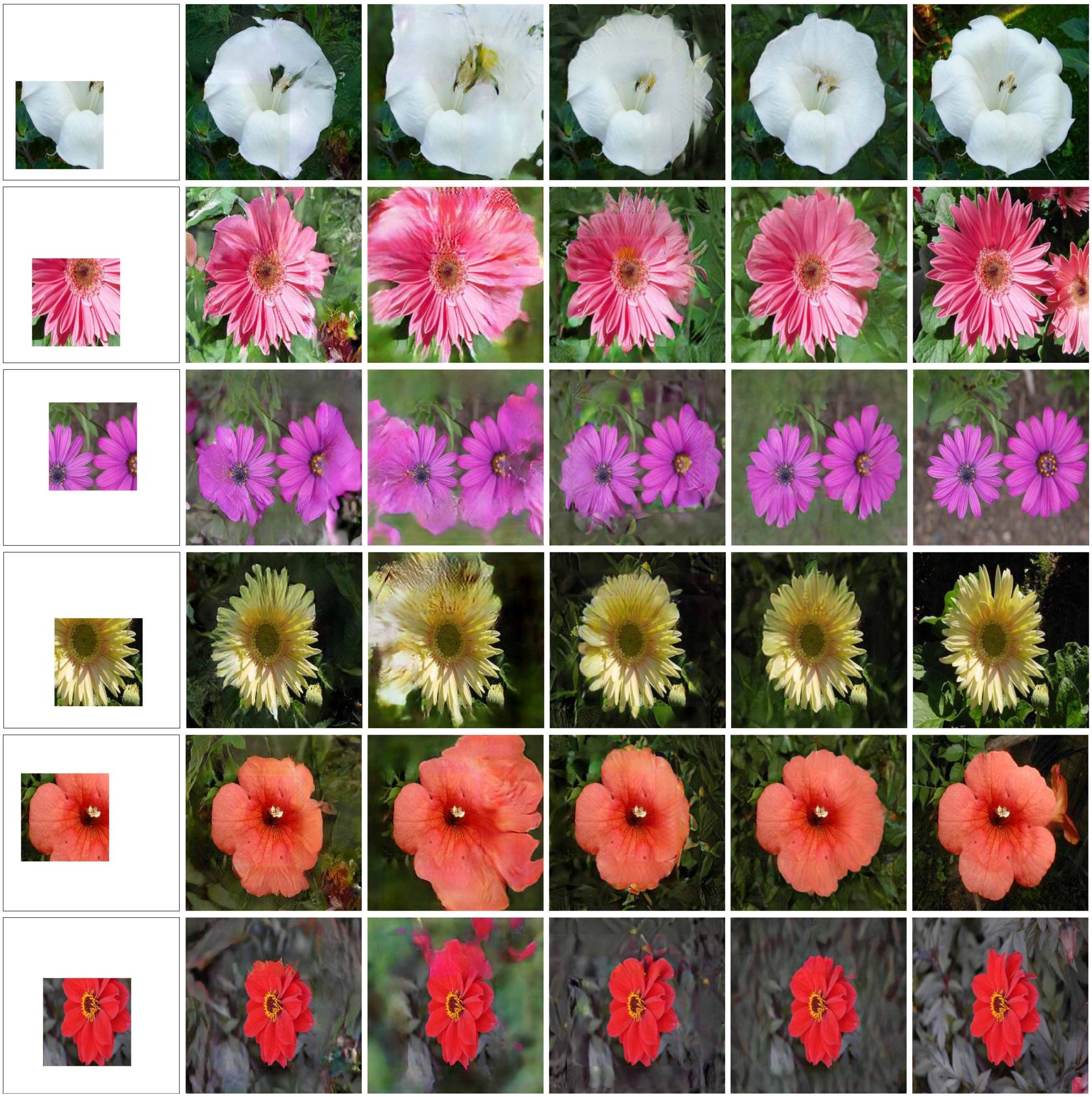


Figure G. Additional qualitative results of regular image outpainting on Flowers [12].

Figure H. Additional qualitative results of regular image outpainting on Paris StreetView [3]. Red boxes mark parts.

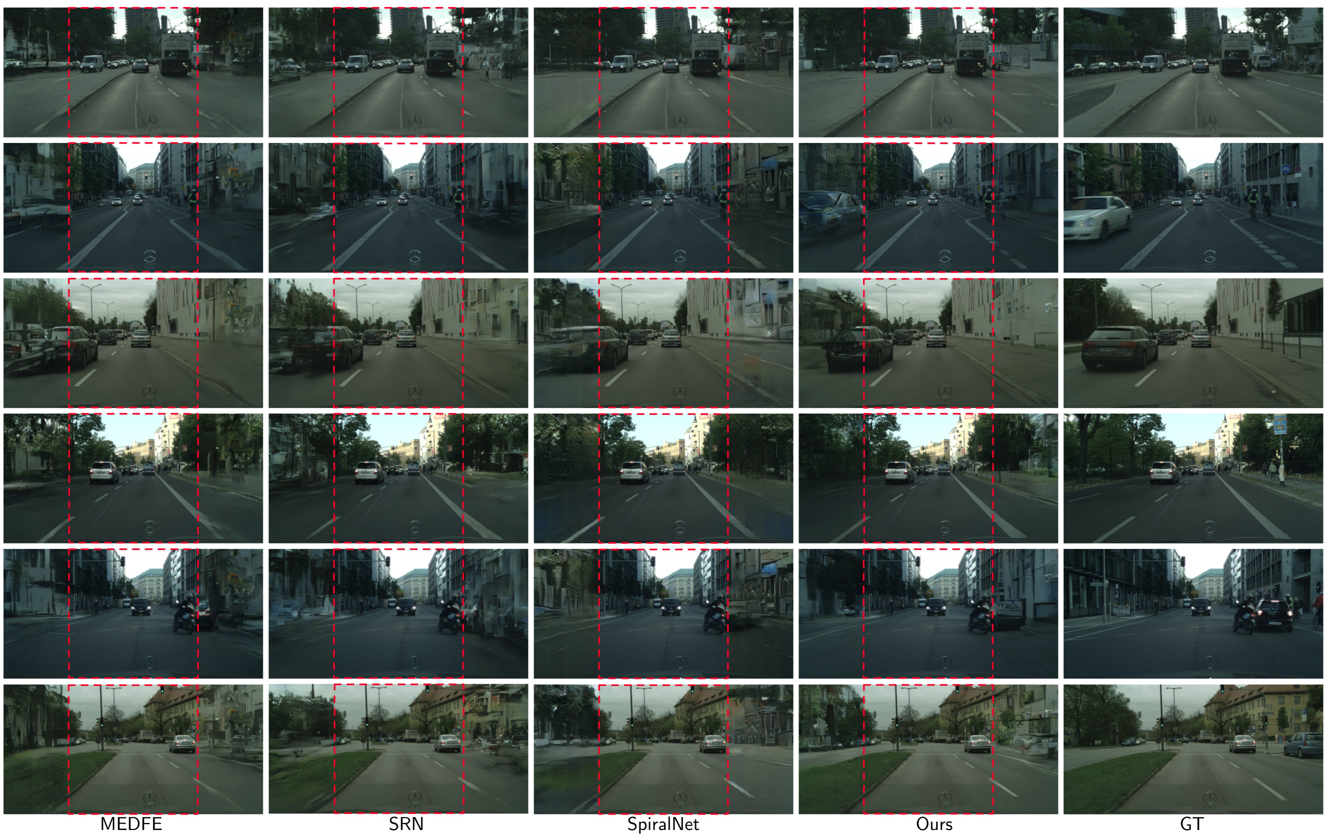


Figure I. Additional qualitative results of regular image outpainting on Cityscapes [2]. Red boxes mark parts.

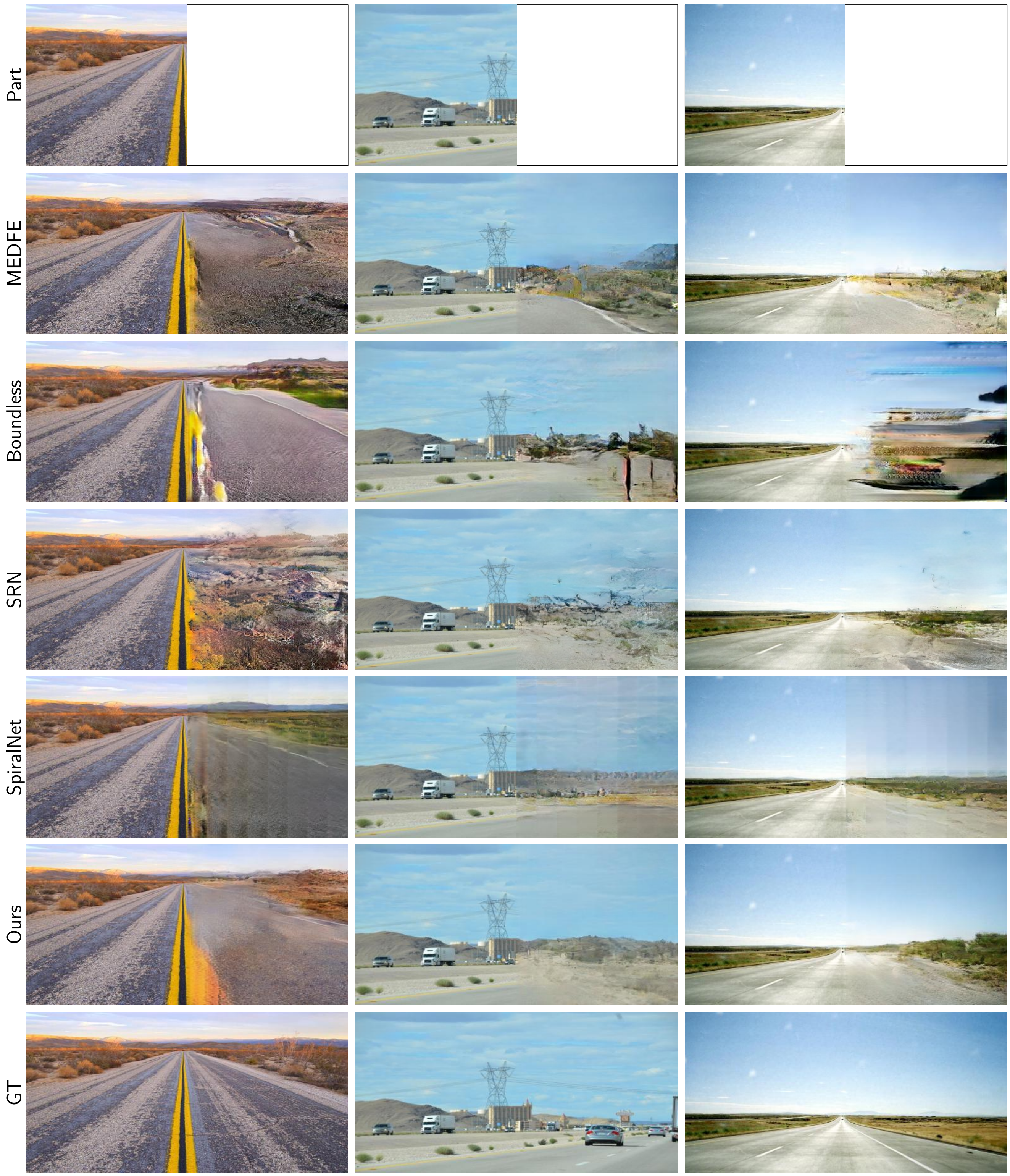


Figure J. Additional qualitative results of regular image outpainting on Places2 Desert Road [15].

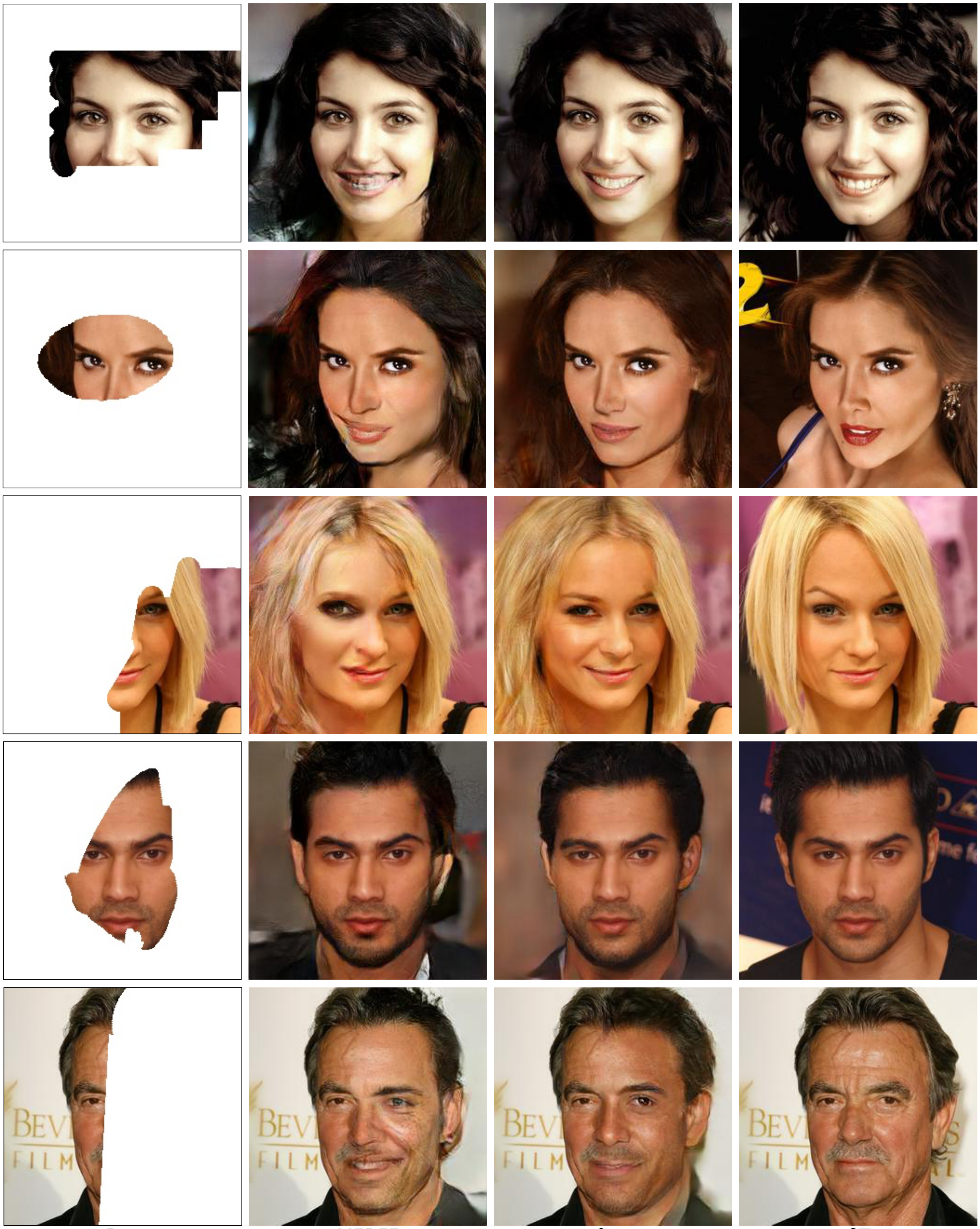


Figure K. Additional qualitative results of irregular image outpainting on CelebA-HQ [7].

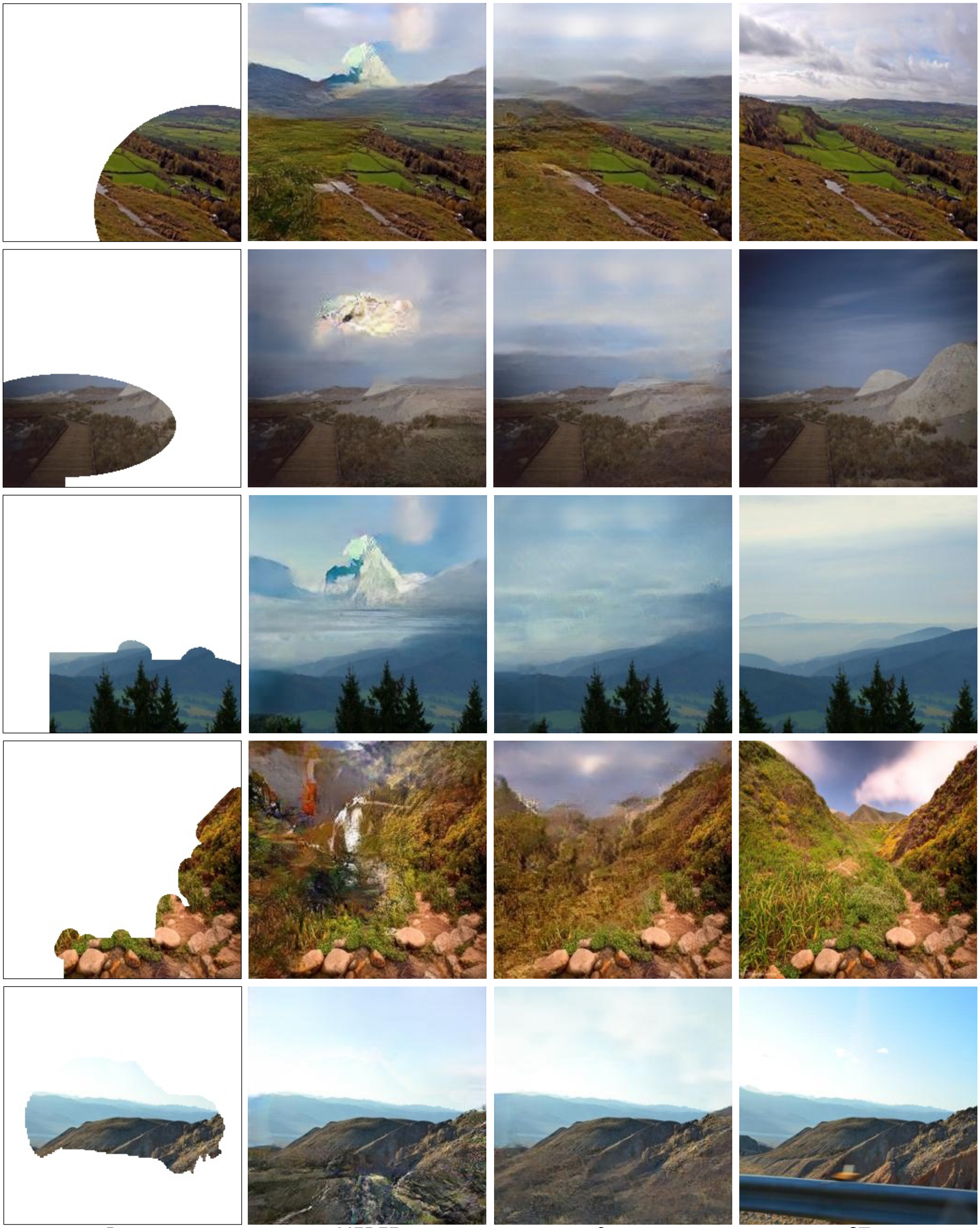


Figure L. Additional qualitative results of irregular image outpainting on Places2 [15].