Painting Outside the Box: Image Outpainting with GANs

Mark Sabini (msabini), Gili Rusak (gili)
CS 230 (Deep Learning), Stanford University

Introduction

- **Image inpainting** is a widely-studied computer vision problem, which involves restoring missing portions within an image
- Current state-of-the-art methods for inpainting involve GANs [1] and CNNs [2]
- We aim to extend [1]'s method for **outpainting,** which extrapolates beyond image boundaries
- Images can then be arbitrarily expanded by recursive outpainting

Problem Statement

- Given an $m \times n$ source image I_s, generate an $m \times (n + 2k)$ image I_o such that
  - I_s appears in the center of I_o
  - I_o looks realistic and natural
- Solve problem for $m = 128, n = 64, k = 32$

Data

Baseline image: 128×128 RGB city image
Dataset: Places365-Standard [3]
- 36,500 256×256 RGB images, downsampled to 128×128
- 100 images held out for validation

Figure 1: **City Image and Places365 Samples**

Data Preprocessing
- Given image I_{tr}, normalize to $[0,1] \rightarrow I_n$
- Define mask M: $M_{ij} = 1 - \mathbf{1}[32 \leq j < 96]$
- Define complement mask $\bar{M} = 1 - M$
- Compute mean pixel intensity μ over I_n
- Set $I_m = \mu M + I_n \odot \bar{M}$
- Stack $I_m \parallel M \rightarrow I_p \in [0,1]^{128\times128\times4}$
- Output (I_n, I_p)

Methods

Training Pipeline
- DCGAN architecture (G, D) used similar to [1]
- Given I_{tr}, preprocess to get I_n, I_p
- Run $G(I_p)$ to get outpainted image I_o
- Run D on I_o and ground truth I_n

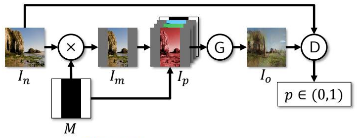


Figure 2: **Training Pipeline**

Training Schedule
- Three-phase training used to condition G, D
- **Phase i:** Optimize loss (i) for T_i iterations using Adam ($\text{lr} = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \varepsilon = 10^{-8}$)
- T_1, T_2, T_3 chosen in $18:2:80$ split
- $\alpha = 0.0004$ controls importance of MSE loss

$$\mathcal{L}_{\text{MSE}}(I_n, I_p) = \|M \odot (G(I_p) - I_n)\|_2^2 \quad (1)$$

$$\mathcal{L}_D(I_n, I_p) = -[\log D(I_n) + \log(1 - D(G(I_p)))] \quad (2)$$

$$\mathcal{L}_G(I_n, I_p) = \mathcal{L}_{\text{MSE}}(I_n, I_p) - \alpha \cdot \log D(G(I_p)) \quad (3)$$

Postprocessing
- I_o renormalized to $[0,255] \rightarrow I_o'$
- I_o' blended with $I_{tr} \odot \bar{M}$ using seamless cloning

Model

Type	f	η	s	n
CONV	5	1	1	64
CONV	3	1	2	128
CONV	3	1	1	256
CONV	3	2	1	256
CONV	3	4	1	256
CONV	3	8	1	256
CONV	3	1	1	256
DECONV	4	1	2	128
CONV	3	1	1	64
OUT	3	1	1	3

Type	f	s	n
CONV	5	2	32
CONV × 4	5	2	64
FC	-	-	512
FC	-	-	1

Figure 3: **Architecture**

Each layer except the last for G (left) and D (right) is followed by ReLU. The output of G, D is followed by sigmoid. Here, η is the dilation factor.

Results

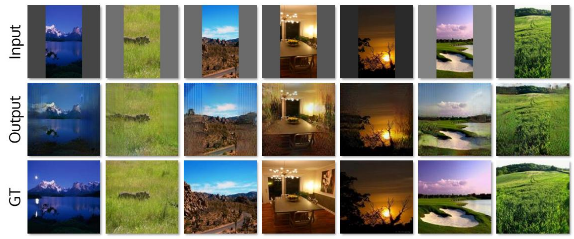


Figure 4: **Outpainting**

Outpainting results for sample of held-out images in validation set, shown alongside original ground truth. Model was trained for 100 epochs (equivalent to 227,500 iterations), using a batch size of 16.

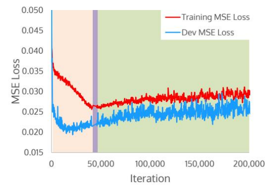
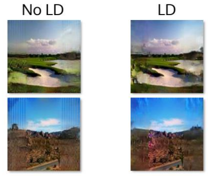


Figure 5: **MSE Loss for Places365**

Training and dev MSE loss on Places365. Phases are illustrated by varying background colors. In Phase 3, the MSE loss increases slightly as we optimize the joint loss (3).

No LD LD



RMSE = 39.68 RMSE = 38.43

Figure 6: **Local Discriminators**

Training with local discriminators (LD) reduced vertical banding and improved color fidelity, but increased artifacts and training time.

Increasing dilation

[1,1,2] [1,2,4] [2,4,8]

η for layers 4, 5, 6 in G

Figure 7: **Effect of Dilation**

The network was trained to overfit on the city image. With insufficient dilation, the network fails to outpaint due to limited receptive field.

GT Outpainted 5×

Figure 8: **Recursive Outpainting**

An outpainted image I_o can be fed as input to the network after expanding and padding. We repeat this recursively, expanding the image's width up to a factor of 3.5. As expected, the noise compounds with successive iterations.

Conclusions

- Image outpainting successfully realized
- Three-phase training aids in stabilizing training
- Dilated convolutions crucial for sufficient neuron receptive field for outpainting
- Recursive outpainting possible, although error and noise compound

Future Work

- Explore sequence models for video outpainting
- Incorporate perceptual and style loss
- Experiment with partial convolutions [2]

References

[1] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (TOG)*, 36(4):107, 2017.

[2] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro. Image inpainting for irregular holes using partial convolutions. *arXiv preprint arXiv:1804.07723*, 2018.

[3] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.