End-to-end Image De-fencing Leveraging Synthetic Data and Adversarial Loss

1st Benjamin Wortman
College of Information Sciences and Technology
Pennsylvania State University
University Park, PA USA
bvw5146@psu.edu

***Abstract*—Image de-fencing is a unique problem in the inpainting community. Unlike standard inpainting problems where the user specifies the area to be corrected, its unreasonable to expect the user to manually segment the entire fence for removal. As such previous work in this field has taken a two stage approach where first either classical or deep learning methods are used to produce a fence mask which is then inpainted to produce a fence free image. In this paper we outline a new end-to-end model trained on synthetic fenced data with adversarial loss to achieve this same result.**

***Index Terms*—De-fencing, inpainting, image de-noising, image aesthetics**

I. Introduction

Although there may be uses for image defencing in other domains, the primary motivation for this problem is in the consumer photography market. A functional image de-fencing model would allow users to clean up any image taken through a fence such as pictures from the zoo, baseball games, hockey games, or anywhere else where there is a barrier between the user and the subject of interest.

Modern methods in GAN based image inpainting can achieve photo-realistic results in inpainting these images, however they require the fenced area to be masked beforehand. Because of this, current state of the art techniques make use of a two stage process for the automatic detection and inpainting of fenced images. Since End-to-end training has been shown to inprove performance on a number of related deep learning tasks, we set out to create an end-to-end model for the simultaneous detection and inpainting of fenced in images.

The contributions of this paper are 2 fold:

- Demonstrate a method for generating large amounts of fenced / ground truth image training pairs.
- Demonstrate an end-to-end model for the simultaneous detection and inpainting of fenced in images

II. Related Work

In the following section we outline related work towards image de-fencing as well as the subtask of image inpainting.

A. Image De-fencing

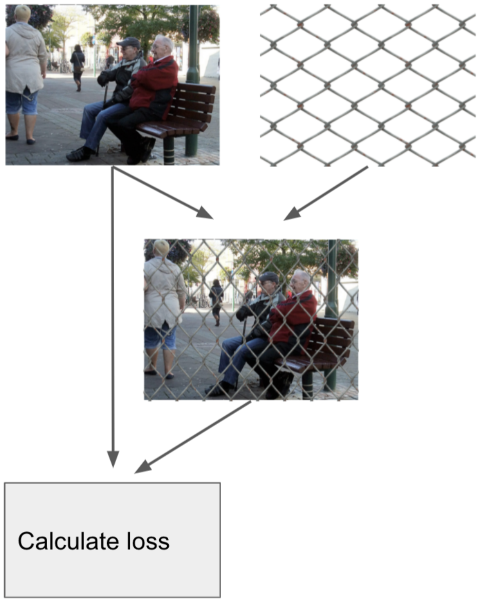

Figure 1. To generate synthetic data, we first combine the raw fence free image with an augmented fence mask. This provides a ground truth and training sample which can be passed through the network in order to calculate the difference between the raw and fence-removed image.

The researchers Liu et al. [1] provided one of the early examples of image defencing. This approach consisted of 3 primary phases. First the lattice is detected using the technique described in [2]. Second, the foreground and background are extracted by clustering based on the standard deviation and colors between pixels in the lattice texels. Finally the image is inpainted using classical techniques as described by Criminisi et al. [3].

Building upon this work, the researchers Khasare et al. [4] show how motion information present in adjacent video frames can be used to identify static fences in the video. One the fence has been identified the area can then be inpainted using information from the adjacent frame. The researchers Jonna et al. [5] took this one step further with the use of CNNs for the detection of the fence before once again inpainting the occluded regions using information from adjacent video frames. For the actual fence detection, the CNN proved much more robust than [2] since a failed detection in one area does not impact the rest of the image.

The researchers Matsui et al. [6] took this one step further by repeating this process for a single image. Once again they first trained a CNN for the segmentation of the fence mask. before applying a simple Gaussian filter to inpaint the image. This pipeline performed incredibly well for standard chainlink fences, however it breaks down for irregular fence shapes and types.

B. Inpainting

Current state of the art in image inpainting almost universally leverages generative techniques such as GANs to produce photorealistic results. In their highly cited paper Pathak et al. [7] demonstrated the use of context encoders with adversarial loss to reconstruct missing areas of images in semantic inpainting tasks. However a major limitation of this approach is its use of fully connected layers as it can't handle images in different resolutions from the original training data. To overcome this, Yang et al. [8] proposed a multi-scale approach which preserves both image content and texture information. More recent work in this field by Xu et al. [9] uses edge information to maintain global structures and further improve results.

III. Methods

A. Data Generation

The largest existing dataset for this problem currently only has roughly 500 training samples with fence masks [10]. However, there is no ground truth beneath these fence masks to score the inpainting results against. To overcome this we proceeded with generating synthetic data by first augmenting these fence masks with color and affine transformations and then overlaying them onto images from the COCO dataset [11]. This process was completed at runtime providing an effectively unlimited amount of training data.

B. Model Training

In defining this problem we can consider the fence in the image as noise which we want to remove from the raw image. We proceeded with autoencoders due to their history in other image de-noising problems. Specifically we proceeded with a RED-Net as described by Mao et al. [12] as the backbone for our model. The intuition behind using a ResNet base is that the skip connections will be able to pass the raw image data for the uncorrupted sections of the image downstream to later layers for reconstruction. The model loss is defined between the original raw image x_r and the synthetic fenced in image x_f as follows:

$$Loss = SSIM(x_r, x_f) + (x_r - x_f)^2 + ADV(x_r, x_f) \quad (1)$$

Here $SSIM$ is the Structural Similarity Index which captures the differences in human perception of the structural information within pairs of images. The second term is standard L2 loss to capture the fine grain details within the reconstruction. We keep this term positive since high L2 loss represents highly dissimilar images. Finally, ADV is adversarial loss from our discriminator network. This was based on an implementation by Radford et al. [13] and included to help produce more photorealistic results.

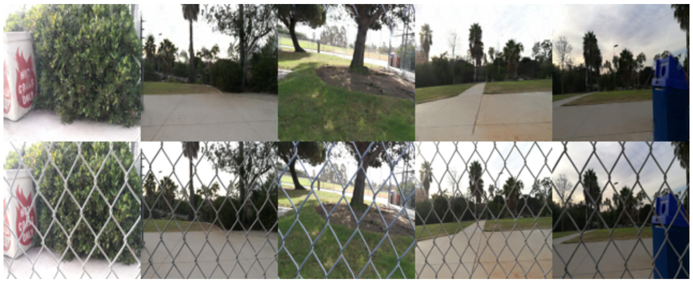

Figure 2. The top row is the output from our model. Below we see the original images.

IV. Results

After training for 100 epochs, this method was able to remove and inpaint the occluded areas in 90 out of 100 testing images as shown in Fig. 2 and Fig. 3. Interestingly our model seemed to struggle with more natural settings and objects in the image. This is possibly due to nature scenes being underrepresented in the COCO dataset which was used for training.

V. Discussion

One benefit of this type of approach is that since modern cell phones have cores optimized for deep learning, this approach could potentially be optimized for live mobile applications. This would allow the user can see in real time how their image will be cleaned up before they even take the photo. Similarly, there may be potential applications in AR/VR where for safety people are required to sit behind a barrier for the event (Baseball, Nascar, Etc.).

A. Limitations

One of the major limitations of this approach is its reliance on pre-gathered fence masks and its inability to generalize to unseen data. If there is a lattice in the image that is not present in the training fences then this model will struggle to identify it. Likewise, if the object being occluded is also not in the training set then the model will struggle to in-paint it. One possible way to overcome this would be to generate random synthetic noise that matches the distributions of the fence masks. Additionally, using existing lattice detection algorithms such as [2] and [4] could potentially be used to gather large amounts of additional fence and lattice data.

In addition to this, to properly training this type of model for the accurate inpainting of high detailed areas such as faces, this requires a considerable investment in GPU time. It may be possible to reduce this training time using transfer learning from models pretrained on larger datasets such as Imagenet. However, this still might not be as efficient as a two stage approach using pretrained models for both fence detection and inpainting.

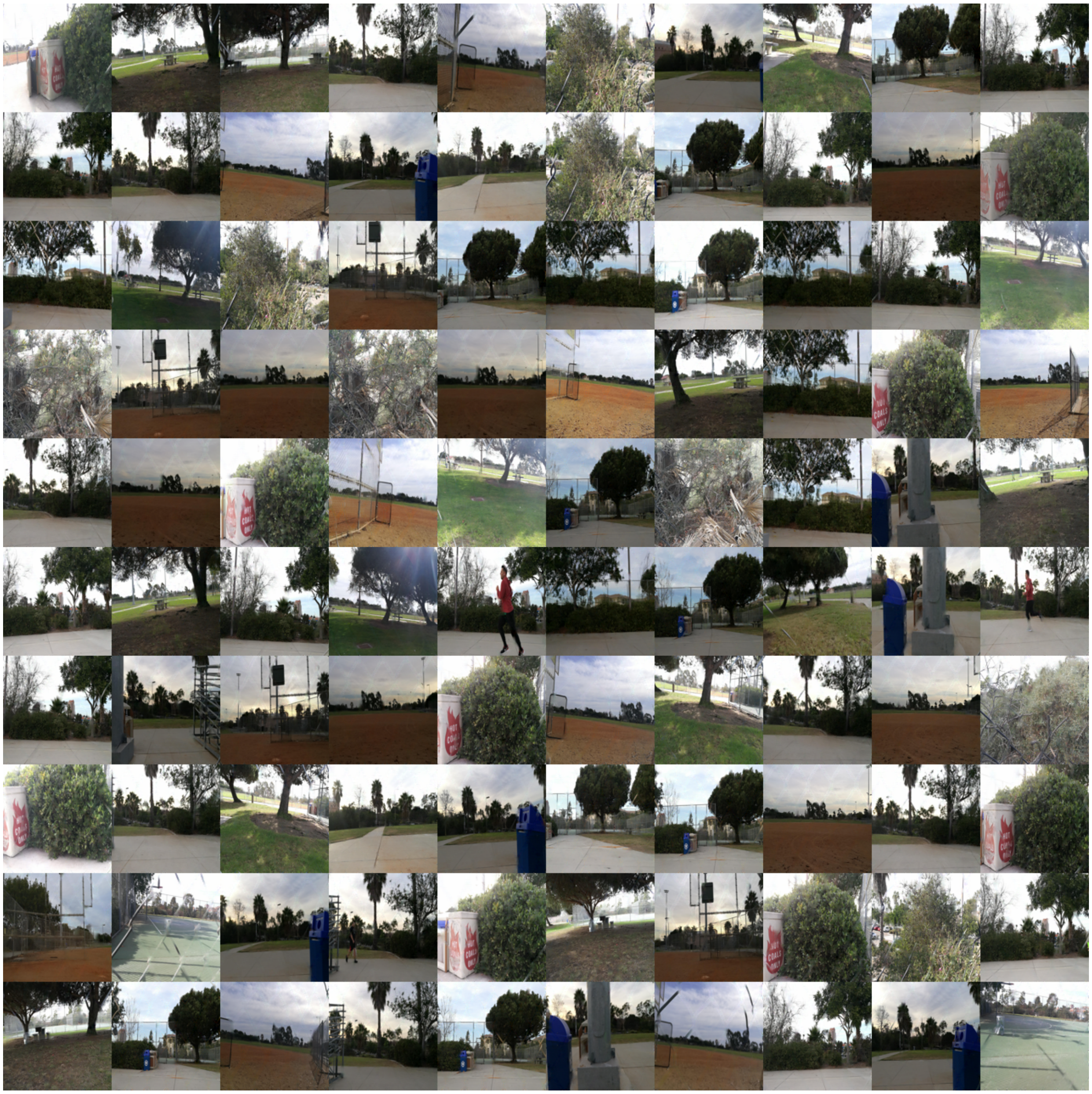

Figure 3. All inpainted images from the test set. Despite only being trained on synthetic data, this method is able to effectively inpaint the large majority of images with photo-realistic results.

VI. Conclusion

This paper outlines and end-to-end method for the simultaneous detection and inpainting of fenced in regions from a single image. Although training on completely synthetic data, the model was able to recognize and removed fences for the large majority of images from the test set. Although the setting for this problems is very specific, this method for synthetic data generation can be applied to any number of inpainting tasks to improve the results.

References

[1] Yanxi Liu, Tamara Belkina, James H. Hays, and Roberto Lublinerman. Image de-fencing. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008. ISSN: 1063-6919.

[2] James Hays, Marius Leordeanu, Alexei A. Efros, and Yanxi Liu. Discovering Texture Regularity as a Higher-Order Correspondence Problem. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, pages 522–535, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[3] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212, September 2004. Conference Name: IEEE Transactions on Image Processing.

[4] Vrushali S Khasare, Rajiv R Sahay, and Mohan S Kankanhalli. Seeing through the fence: Image de-fencing using a video sequence. In *2013 IEEE International Conference on Image Processing*, pages 1351–1355, September 2013. ISSN: 2381-8549.

[5] Sankaraganesh Jonna, Krishna K. Nakka, and Rajiv R. Sahay. My camera can see through fences: A deep learning approach for image de-fencing. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 261–265, November 2015. ISSN: 2327-0985.

[6] Takuro Matsui and Masaaki Ikehara. Single-Image Fence Removal Using Deep Convolutional Neural Network. *IEEE Access*, 8:38846–38854, 2020. Conference Name: IEEE Access.

[7] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context Encoders: Feature Learning by Inpainting. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2536–2544, June 2016. ISSN: 1063-6919.

[8] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. High-Resolution Image Inpainting Using Multi-scale Neural Patch Synthesis. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4076–4084, Honolulu, HI, July 2017. IEEE.

[9] Shunxin Xu, Dong Liu, and Zhiwei Xiong. E2I: Generative Inpainting From Edge to Image. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(4):1308–1322, April 2021.

[10] Chen Du, Byeongkeun Kang, Zheng Xu, Ji Dai, and Truong Nguyen. Accurate and Efficient Video De-Fencing Using Convolutional Neural Networks and Temporal Information. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, July 2018. ISSN: 1945-788X.

[11] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. *CoRR*, abs/1405.0312, 2014. arXiv: 1405.0312.

[12] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[13] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434 [cs]*, January 2016. arXiv: 1511.06434.