Photobombing Photobombers

Mihir Rana, Thibault Févry, and Kenil Tanna

Center for Data Science, New York University

Objectives

The goal of our project is to propose a pipeline that can efficiently remove photobombers in images and replace them with background.

Introduction

Computer vision research is very prolific on the topic of image segmentation, with recent SOTA methods achieving very high performance. More recently, in the wake of GANs introduced by I. Goodfellow, researchers have explored image generation. Our goal is to leverage these two approaches to replace photobombers in still images while leaving as few artefacts as possible. More generally, our approach should be applicable to quickly remove an object in an image and replace it with background, which has many implications in image editing, movie post-production and many other domains. This proof-of-concept also shows how easily images can be manipulated. We discuss these ethical implications in further detail in the report.

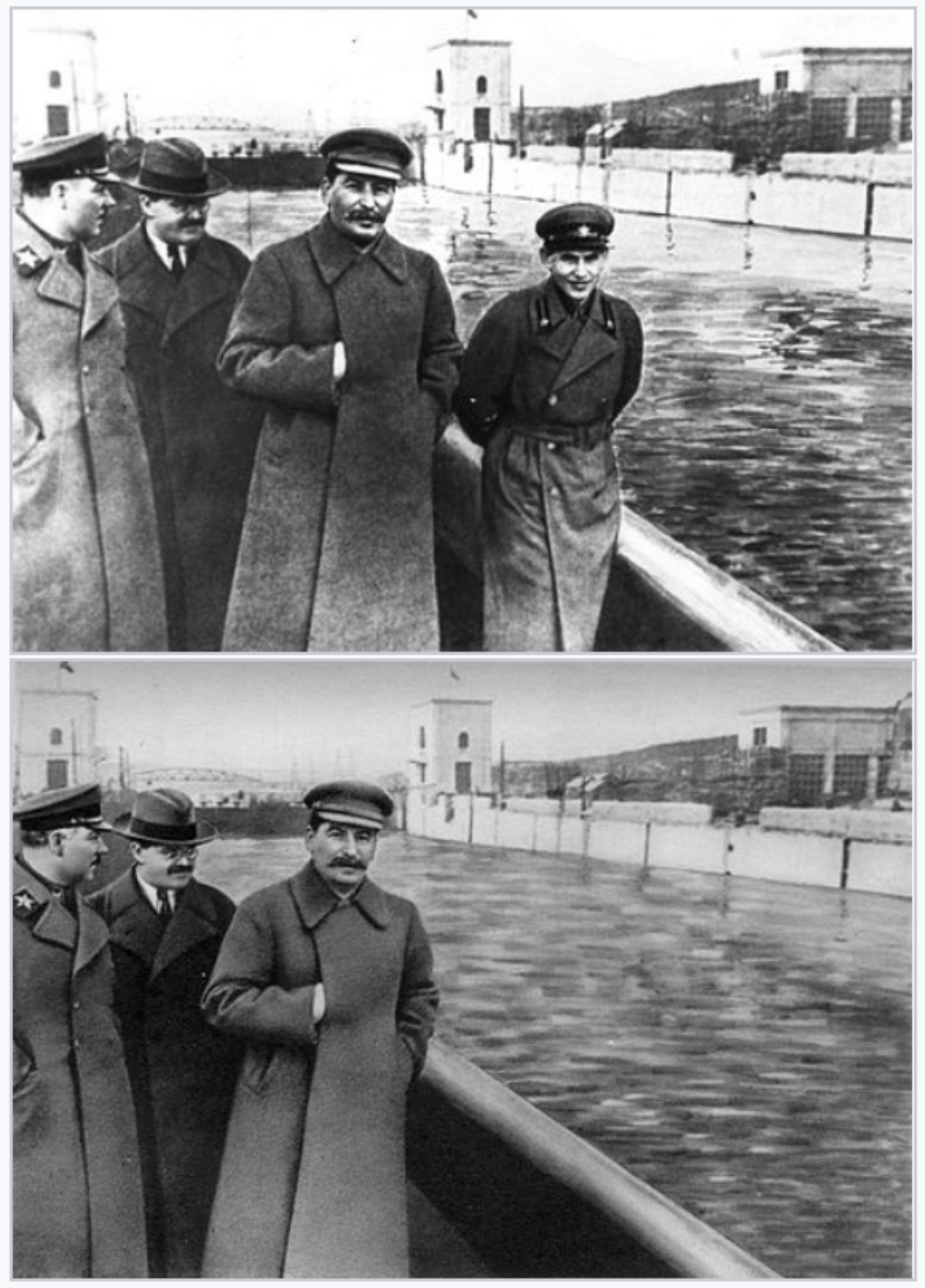

Figure 1: Early application of our method: removal of Nikolai Yezhov, executed in 1940, by USSR's censorship service

Proposed Pipeline

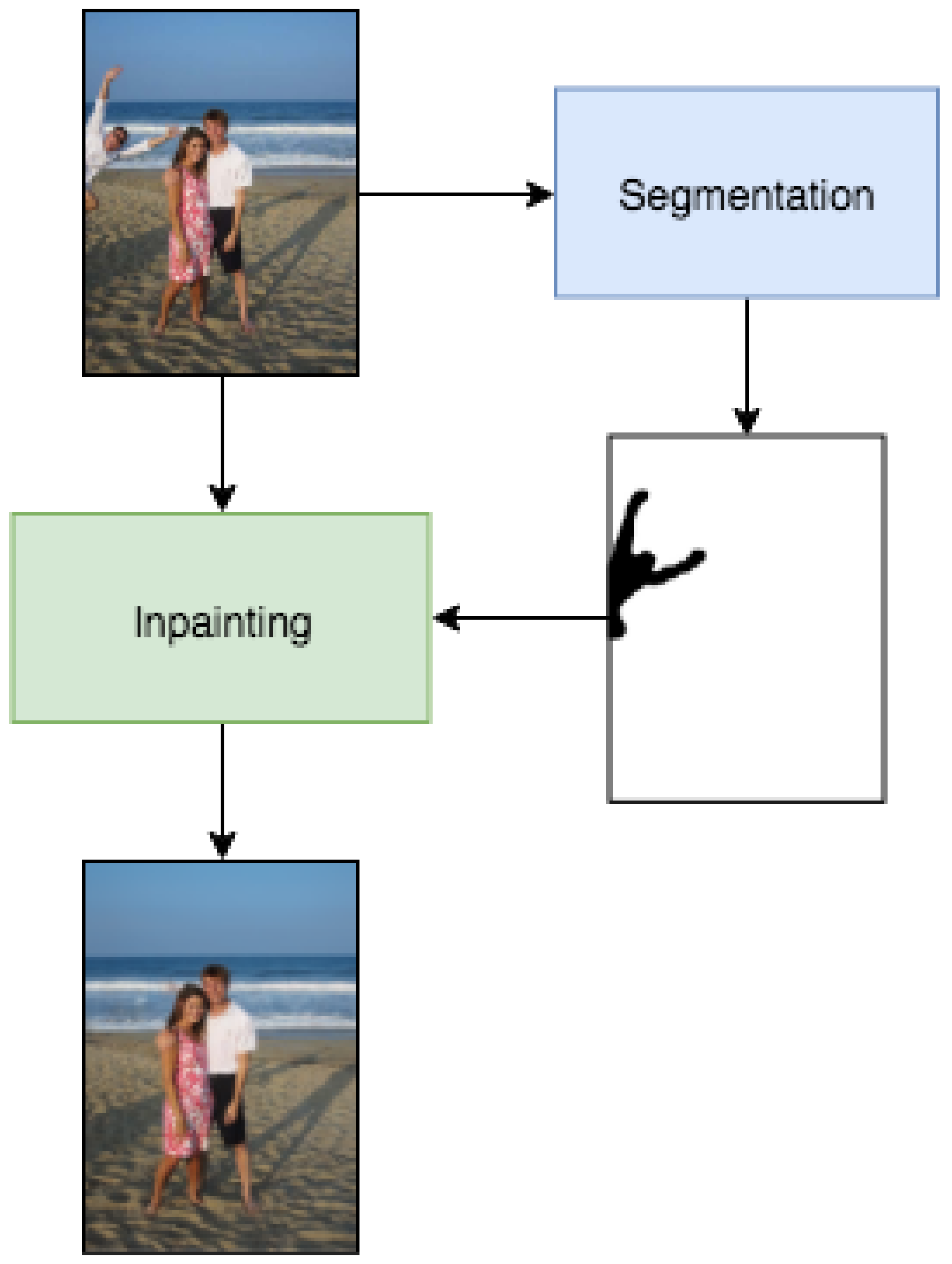

Figure 2: Pipeline used to remove photobombers

Segmentation Methods

We use Mask R-CNN [1] for segmentation. Given a pixel, we find the associated mask and remove it.

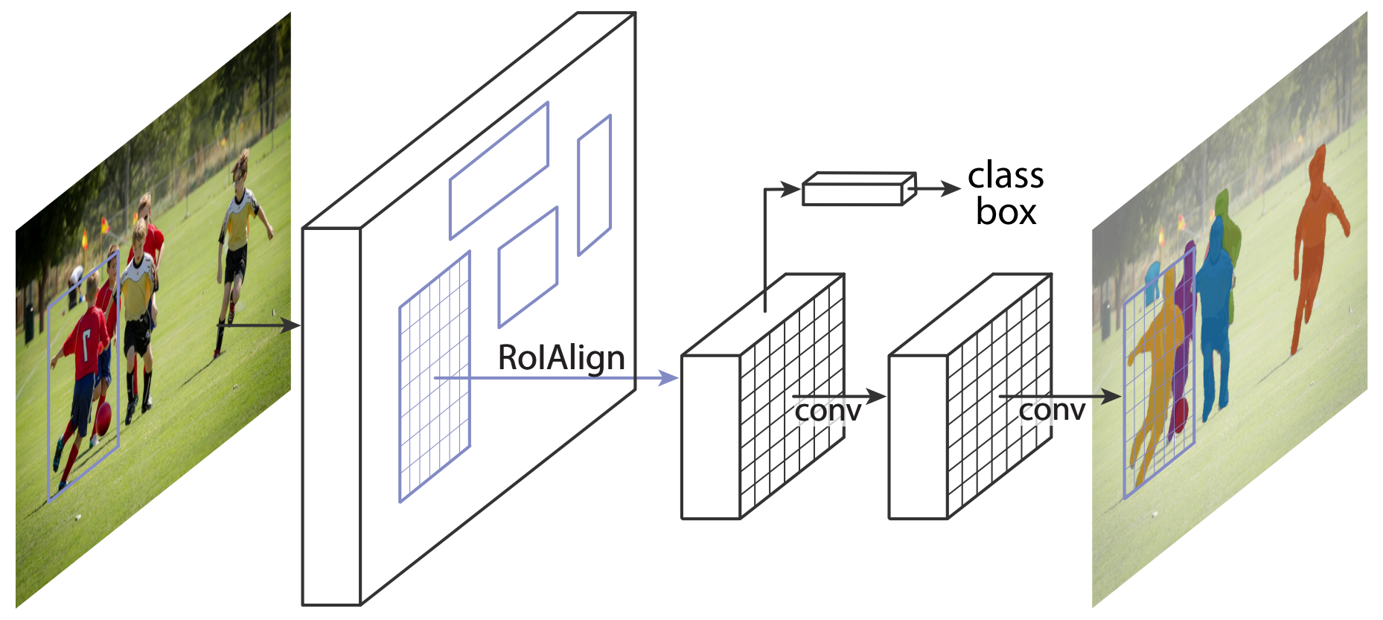

Figure 3: **Mask R-CNN** architecture (image source)

We use the Detectron implementation. Our design consideration criteria were:

- Segmentation quality: It provides high-quality masks. In particular, we found it strongly outperforms other models on overlapping objects.
- Speed: Mask R-CNN runs at > 1 fps in our experience (5 in paper, but on better hardware).
- 81 pre-trained classes: from COCO), many of which relevant (such as person, dog, car or bird).

Inpainting Methods

We compare two methods.

1. Deep Image Prior [2] It is an unsupervised denoising auto-encoder. To train, initialize noise z, network f_θ (see below) with random weights and optimize $\min_\theta ||(f_\theta(z) - x_0) \odot m||^2$ where x_0 is the masked out image and m the mask.

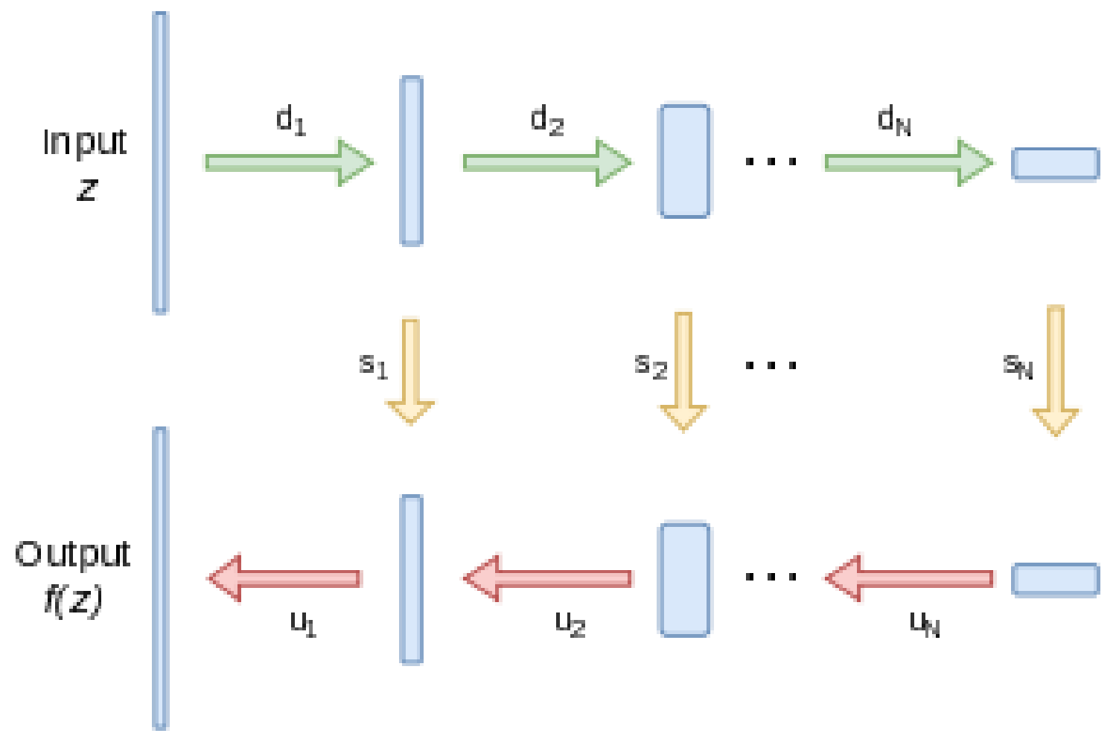

Figure 4: Architecture of Deep Image Prior

2. WGANs with contextual attention [3], a supervised method trained on ImageNet. Tricks and contextual attention make training efficient.

Results

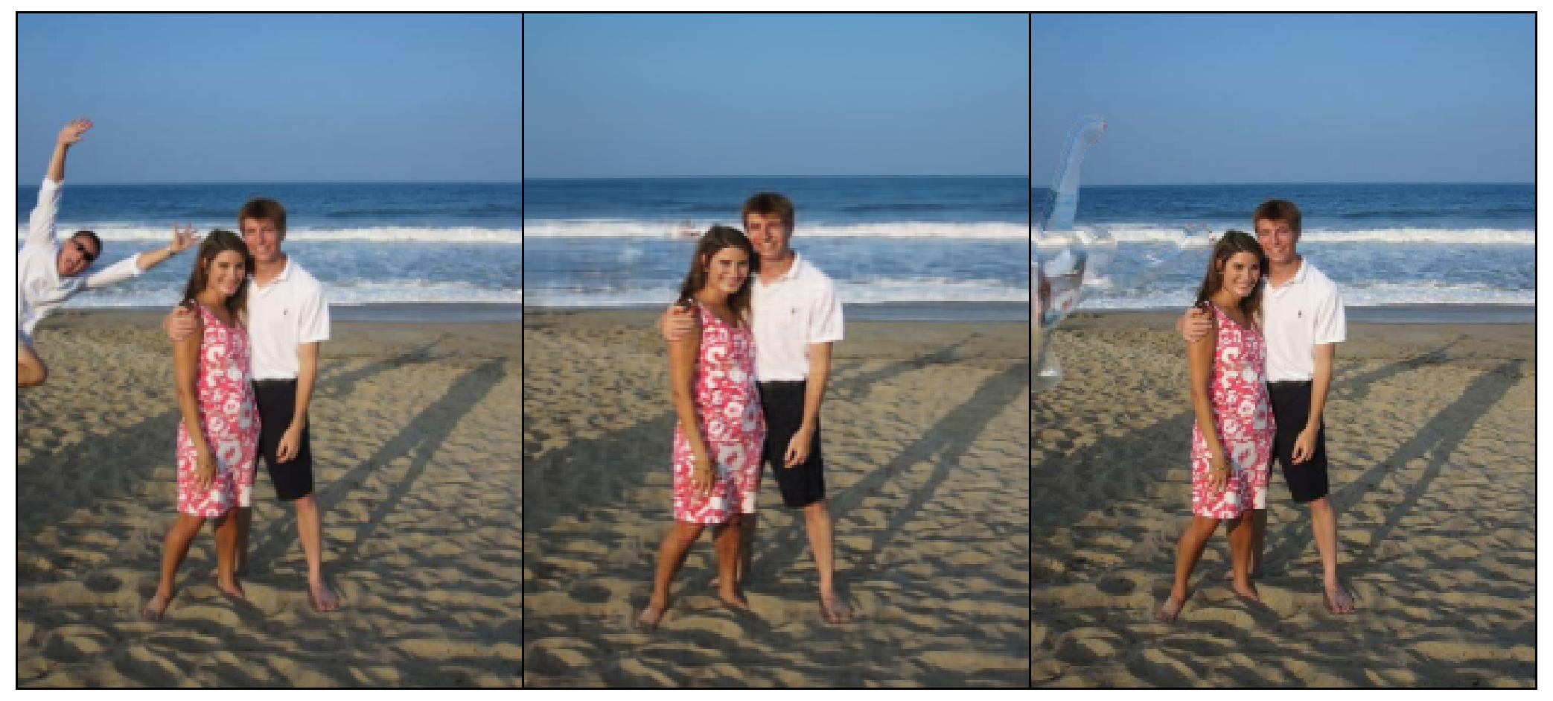

Figure 5: Original image (left), DIP Result (middle), GIICA Result (right)

- DIP (middle) outperforms GIICA (right) although it is unsupervised
- Deeper encoder-decoder architectures with skip connections were found to outperform others
- Poor performance on selfies and images with little background
- DIP starts overfitting if too many iterations; 5000 generally works best. Takes ~5m on a P40.

Results(cont.)

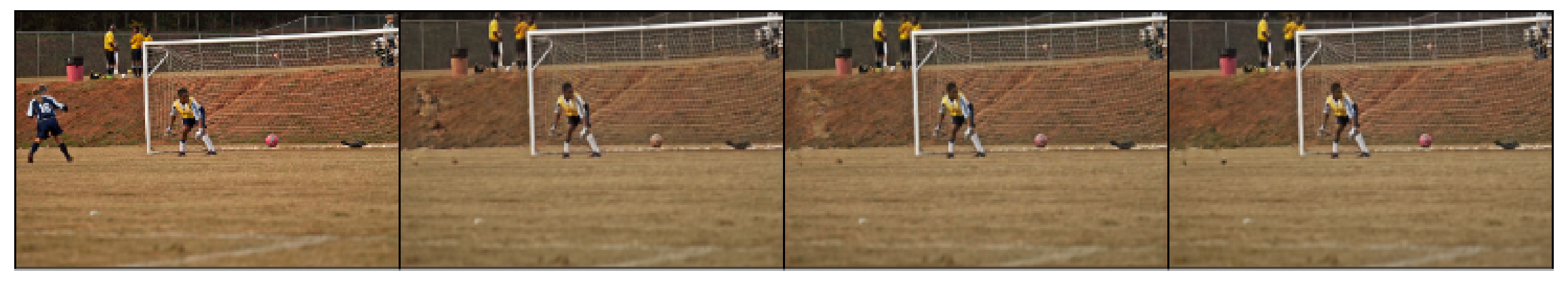

Figure 6: Generated images from U-Net, ResNet and encoder-decoder architecture with skip connections

Conclusion

We propose a pipeline to replace photobombers in still images. We compare different architectures and show Deep Image Prior works best in general. The generated images show few artefacts.

Future Work

In the future, we hope to investigate the following:

- Using weakly supervised methods to remove any object (based on [4])
- Speeding up inpainting so that the full demo can be integrated on a website
- Automatically detecting photobombs in large data sets (if we build a data set from social media / our website)
- Automatically detecting optimal architecture for inpainting based on image context and masks
- Detecting masks associated with an object (dog leash, shadows, etc.) during segmentation and inpaint them

References

[1] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," *CoRR*, vol. abs/1703.06870, 2017.

[2] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Deep image prior," *CoRR*, vol. abs/1711.10925, 2017.

[3] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," *CoRR*, vol. abs/1801.07892, 2018.

[4] R. Hu, P. Dollár, K. He, T. Darrell, and R. Girshick, "Learning to segment every thing," *arXiv preprint arXiv:1711.10370*, 2017.