Project Proposal

Nik Bhatt UNI: nb3097

What I want to study

My midterm paper opened my eyes to the state of art for inpainting algorithms, particularly for image synthesis. That said, I wonder about the general applicability of these newer algorithms. Each research team uses the same datasets, looking for small improvements in aggregate, and showing cherry-picked special cases. For example, Learned Perceptual Image Patch Similarity (LPIPS) is one of the key quality metrics in inpainting research. Reviewing various papers, I noticed that there typically only a small overall improvement from paper to paper. No variety in the datasets, combined with similar results, might indicate that the solutions or approaches are highly tuned to one problem domain and lack generality. Further, while image synthesis is intriguing, I'm personally interested in the type of inpainting that my customers (photographers) would want - small defect repair and object removal. Of those two, object removal is much more difficult, so I will target that use case.

During my literature review, I read a paper by Xiang, et al. [1] that tested many different algorithms with the same data sets in an attempt to provide a consistent basis of comparison. I would like to take that idea further, to see how well the latest algorithms perform in the object removal use case, using a completely unseen dataset. I would also include an earlier algorithm as a form of "control" (HiFill, in the chart above, also tested in [1] and in my midterm project).

Background

The Xiang paper covers inpainting in general, but its primary contribution is its analysis of the performance of many algorithms against a consistent collection of datasets. It measures multiple attributes including model size, inpainting quality, and training time. Since that study was published, newer GANs have emerged. These seem to be focused on image synthesis rather than on classic inpainting problems (repair and object removal). While these models are promoted as addressing these cases, the papers rarely discuss how well they solve them (nor does Xiang). As a result, I noticed multiple opportunities to add novelty through my own tests and analysis.

Opportunities for novelty

I have identified four main areas that are advancements over the Xiang paper:

	Xiang	Mine	Benefit
Datasets	ImageNet, CelebA, CelebA-HQ Places, and Paris StreetView. (These are the same datasets that the models are trained against.)	OpenImages v7 (https://storage.googleapis.com/openimages/web/download_v7.html)	OpenImages is a completely novel dataset to these models, eliminating the chance of contamination or overfitting from the training process.
Models	A wide range of models up to 2021	Newer models up to 2023	Only one model (the "control") is analyzed in the other paper.
Masks	NVidia's Irregular Mask set, which is a random mask generator.	Object Detection masks generated by Facebook's Detectron2.	These masks represent what end users would want to use an inpainting algorithm for.

	Xiang	Mine	Benefit
Image Size	128 x 128 256 x 256 512 x 512 (a few times)	256 x 256 512 x 512 1024 x 1024 2048 x 2048 (if possible)	Larger images can expose algorithms that do not scale well or cannot run on consumer-grade hardware.

I would compute three metrics: Fréchet Inception Distance (FID), Learned Perceptual Image Patch Similarity (LPIPS), and Structural Similarity (SSIM). These are the primary metrics used in current inpainting research. I would start by replicating the findings in the papers, using their pre-trained models. Then I would switch to the new dataset and masks. To enable these models to run more quickly, I plan to convert each of them to Apple's CoreML format which allows them to be run on Apple's Neural Engine, rather than on the CPU.

Ablation

1. I can use the same NVidia irregular mask sets to see how much of the differences I find in quality are due to the masks chosen vs. the image set.
2. I might also use the same datasets as Xiang, but that may not possible in the time I have.

Research Questions

1. Are newer GANs able to remove objects with good quality results, or has that been diminished in the race for image synthesis? Are they better than the control?
2. Can I reproduce the results as described in the papers?
3. How different are the reproduced metrics vs. ones using the OpenImages dataset?
4. What does the difference tell us? If the OpenImages results are worse, does that imply overfitting / overtraining? Is there an explanation if the OpenImages results are better?
5. Which recent algorithms are best at object removal with an unseen dataset, both in terms of quality and runtime? How well do they handle larger images on a standard laptop?

Target User Community

Researchers and product development engineers interested in inpainting for object removal.

Why would users use your project?

My approach uses novel images and mask generation. Users could use these techniques to assess how well future inpainting algorithms handle foreign data sets for object removal. Researchers could use this dataset and mask generation to train or evaluate their models.

References:

[1] Xiang, et al. Deep learning for image inpainting: A survey. In *Pattern Recognition, Volume 134, Feb 2023.* https://www.sciencedirect.com/science/article/abs/pii/S003132032200526X

[2] Sargsyan, et al. MI-GAN: A Simple Baseline for Image Inpainting on Mobile Devices. In *International Conference on Computer Vision* (ICCV), 2023. https://openaccess.thecvf.com/content/ICCV2023/papers/Sargsyan_MI-GAN_A_Simple_Baseline_for_Image_Inpainting_on_Mobile_Devices_ICCV_2023_paper.pdf