Guided Image Inpainting: Replacing an Image Region by Pulling Content from Another Image

Yinan Zhao* Brian Price+ Scott Cohen+ Danna Gurari*

*University of Texas at Austin +Adobe Research

Motivation

Image Inpainting: the task of filling in the lost part of an image.

Previous Works

- When synthesizing missing image regions conditioned on surrounding context, users cannot control what content is synthesized (for example, Context Encoder[1]).
- When cutting, pasting, and blending a semantically similar patch from another image, results can be unrealistic when the pasted content differs from the context of the image.

Incomplete | Context Encoder[1] | Cut-Paste-Blend[2]

Our Idea: use another image to "guide" the synthesis process within a deep learning framework (**Guided Image Inpainting**).

Incomplete | Guidance | Ours

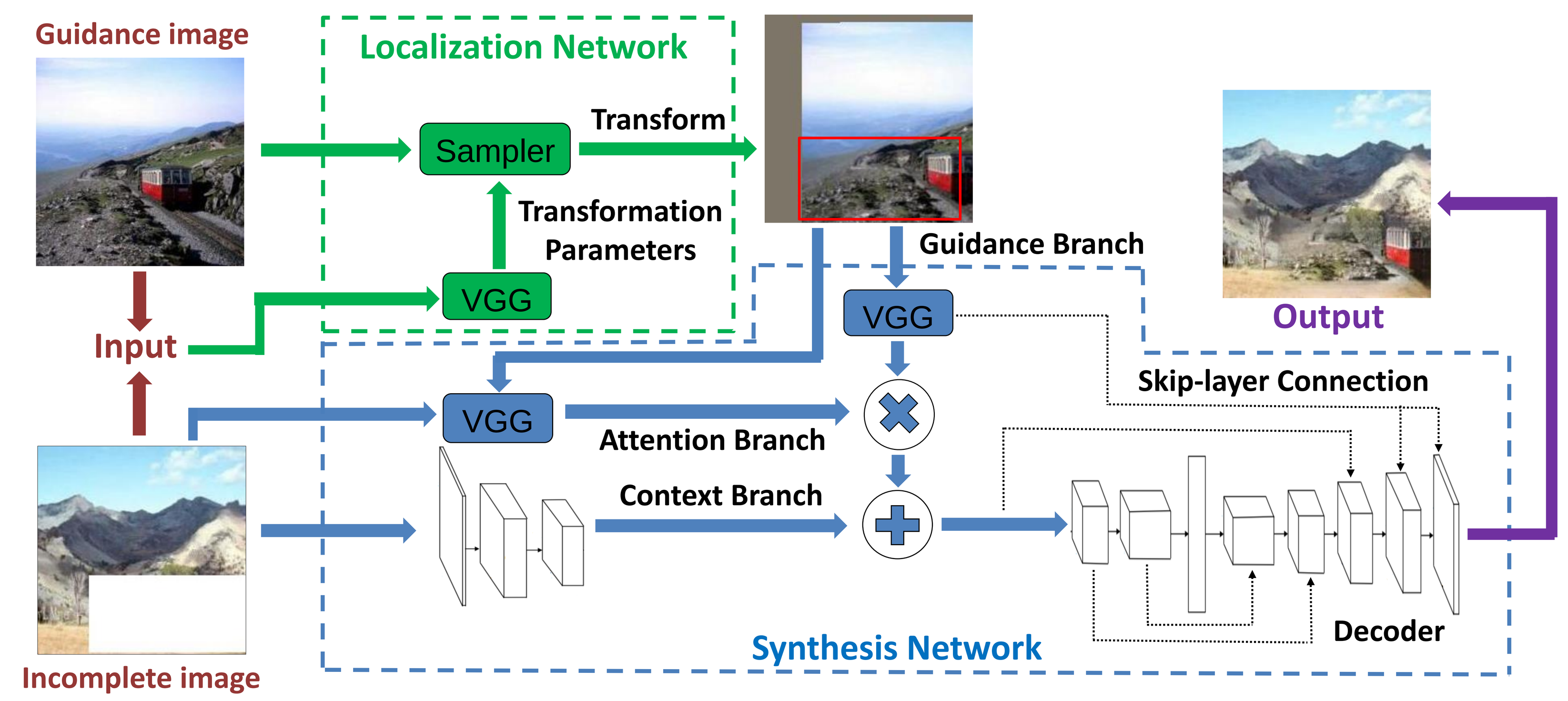
Advantage: can synthesize diverse realistic hole-fillings and users can control the content to use

Key Challenge: identify inconsistent regions between the guidance patch and surrounding context, and then synthesize new content to match those regions with the context.

Approach

Architecture

Given an *incomplete image* and *guidance image*, Our model identifies a patch in the guidance image to replace the hole, and synthesizes new content to fit within the image context, informed by the identified patch.



Localization Network identifies a patch from the guidance image to inform the synthesis process, and aligns the guidance patch with the hole.

Synthesis Network encodes the guidance patch and incomplete image, locates inconsistent regions between the patch and surround image context, and synthesizes new content to clean inconsistencies and fill the hole.

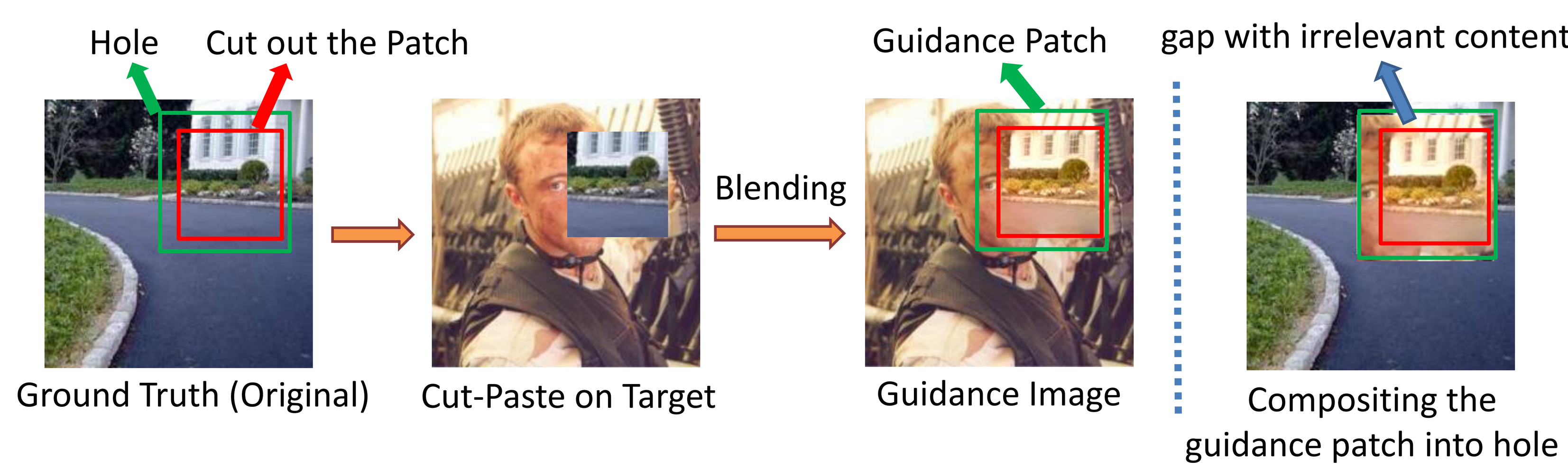
Training Data

Goal: Large-scale data is needed to train the model.

Challenge: There is no clear way to generate the ground truth.

Solution: train the model to reconstruct the original patch given a corrupted one as guidance. Then we can use the original patch as ground truth.

Key Question: how well does the model trained on the synthetic images generalize to real images?



We create a synthetic dataset to train the model and demonstrate that it generalizes well to real images in the experiments.

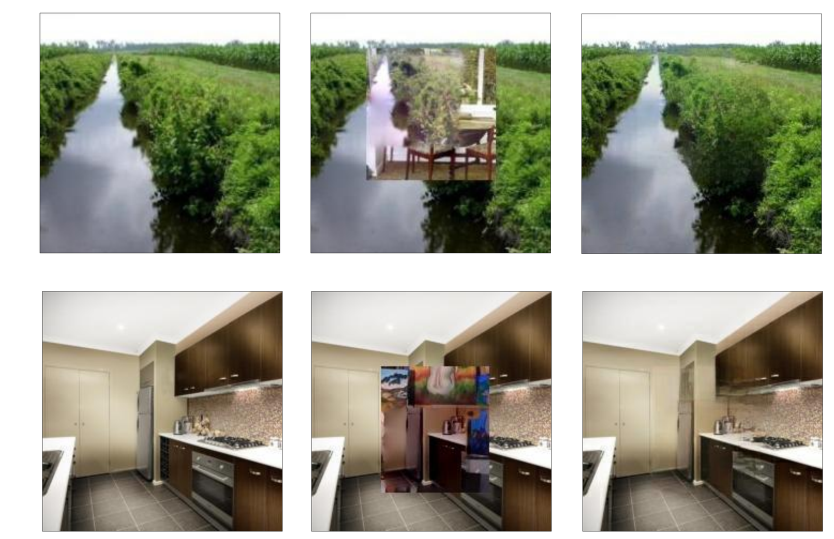
Evaluation

Image Restoration

Our method outperforms all the baselines by at least 5.34% in Mean L1 Loss, 1.93% in Mean L2 Loss, and 4.13dB in PSNR.

Method	Mean L1	Mean L2	PSNR
CAF	15.43%	5.09%	14.38dB
CE	12.91%	3.21%	15.91dB
HR	13.05%	3.29%	15.83dB
GLCIC	13.28%	3.47%	15.56dB
PB	13.63%	3.28%	15.41dB
IM	12.23%	3.04%	16.55dB
DH	18.87%	6.02%	12.73dB
Ours	**6.89%**	**1.11%**	**20.68dB**

Original | Composite | Ours



Absolute Realism

33% and 36% of our synthesized images are deemed to be real by human reviewers with Retrieval (a) and (b) respectively, outperforming all the baselines.

Method	NI	CE	HR	CAF	PB	DH	GLCIC	IM	Ours
Retrieval (a)	97.7%	10.0%	14.0%	31.0%	18.0%	23.0%	14.0%	23.0%	**33.0%**
Retrieval (b)	97.7%	22.0%	15.0%	16.0%	20.0%	22.0%	12.0%	27.0%	**36.0%**

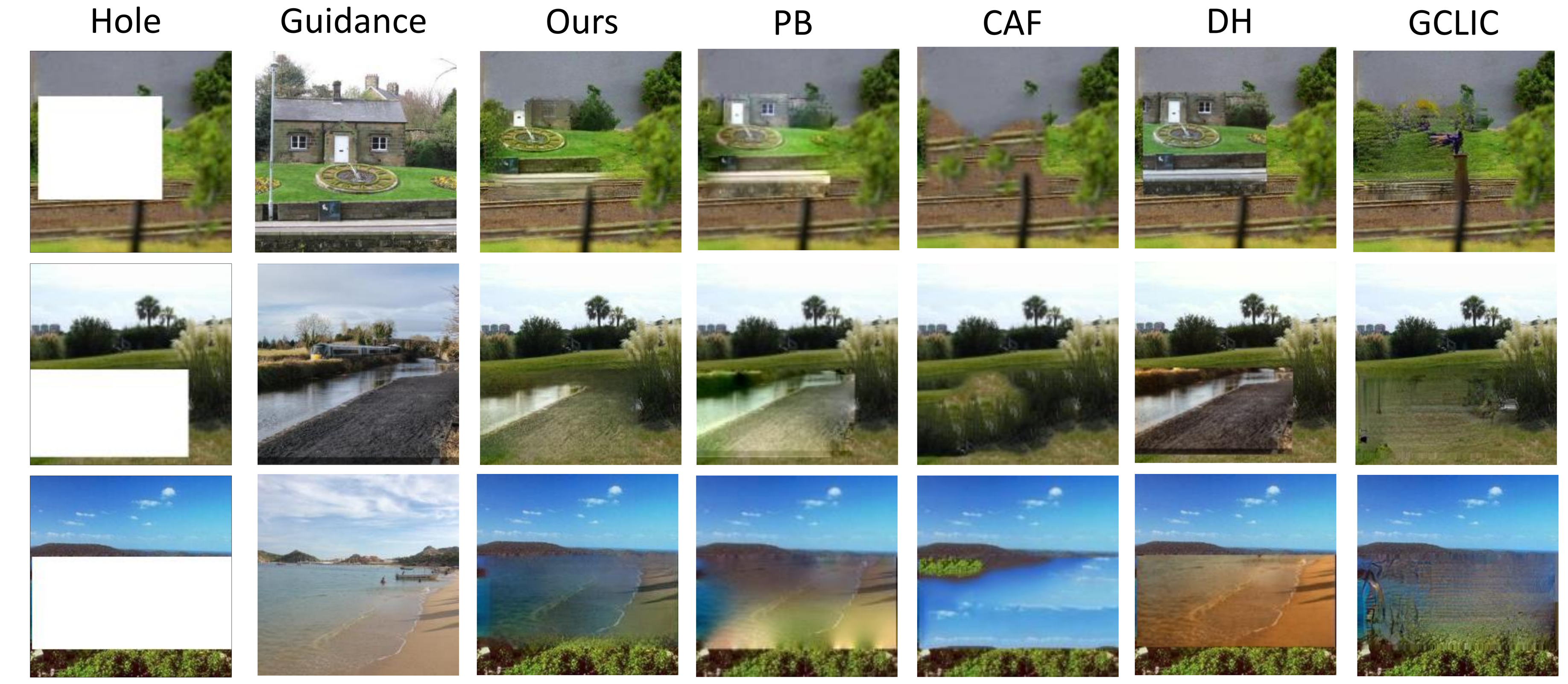
Retrieval (a) uses the full original image to retrieve the guidance image while Retrieval (b) uses the incomplete image with its hole filled by CAF to retrive the guidance image.

Relative Realism

People rate our synthesized images more realistic than all baselines for at least 66% of test images.

Method	HR	PB	DH	CAF	CE	IM	GLCIC
Retrieval (a)	76%	76%	71%	70%	70%	67%	66%
Retrieval (b)	71%	73%	72%	70%	73%	67%	70%

Hole | Guidance | Ours | PB | CAF | DH | GCLIC



[1] Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." *CVPR* 2016.
[2] Pérez, Patrick, Michel Gangnet, and Andrew Blake. "Poisson image editing." TOG 2003