Inpainting Applied to Facade Images: A Comparison of Algorithms

Hochschule Niederrhein
University of Applied Sciences
iPattern
Institut für Mustererkennung
Institute for Pattern Recognition

Willy Fritzsche, Steffen Goebbels, Simon Hensel, Marco Russinski, and Nils Schuch

{willy.fritzsche, marco.russinski, nils.schuch}@stud.hn.de, {steffen.goebbels, simon.hensel}@hsnr.de

INTRODUCTION

Many cities provide a textured 3D city model for planning and simulation purposes. Usually, the textures are automatically taken from oblique aerial images showing occlusions by, e.g., vegetation. These objects have to be segmented and then removed from facade textures. In this study, we investigate the ability of different non-specialized inpainting algorithms to continue facade patterns in occluded facade areas. In particular, very useful results are obtained with the neural network "DeepFill v2" trained with transfer learning on freely available facade datasets and the "Shift-Map" algorithm.

ALGORITHMS

We tested six algorithms, including two standard algorithms from the OpenCV library, two algorithms from the xphoto package of OpenCV, and two deep neural networks.

Local, Diffusion-based Inpainting Algorithms of OpenCV

Both **Navier-Stokes** and **Telea algorithms** continue the patterns inward from the boundary of the occluded regions.

Global Inpainting/Texture Synthesis Algorithms from the Xphoto Module

A shift-map consists of offsets that describe how pixels are moved (shifted, transformed) from a source to a target image region. By choosing an occluded area as a source, the **shift map algorithm** computes an optimized shift-map to do example-based inpainting.

Frequency Selective Reconstruction (FSR) uses Fourier analysis to reconstruct the missing pixels. Unfortunately, the current implementation of FSR has some problems with large images, so we had to scale them down in order to get results.

DeepFill V2 Algorithm

The "Free-Form Image Inpainting with Gated Convolution" [2] network is based on gated convolution. This allows the network to learn how to apply convolution kernels to incomplete data including a mechanism for dynamic feature selection. It consists of three subnets: a network for coarse inpainting, a contextual attention network for adding details, and a third network which computes an adversarial loss that is linearly combined with an l_1 loss.

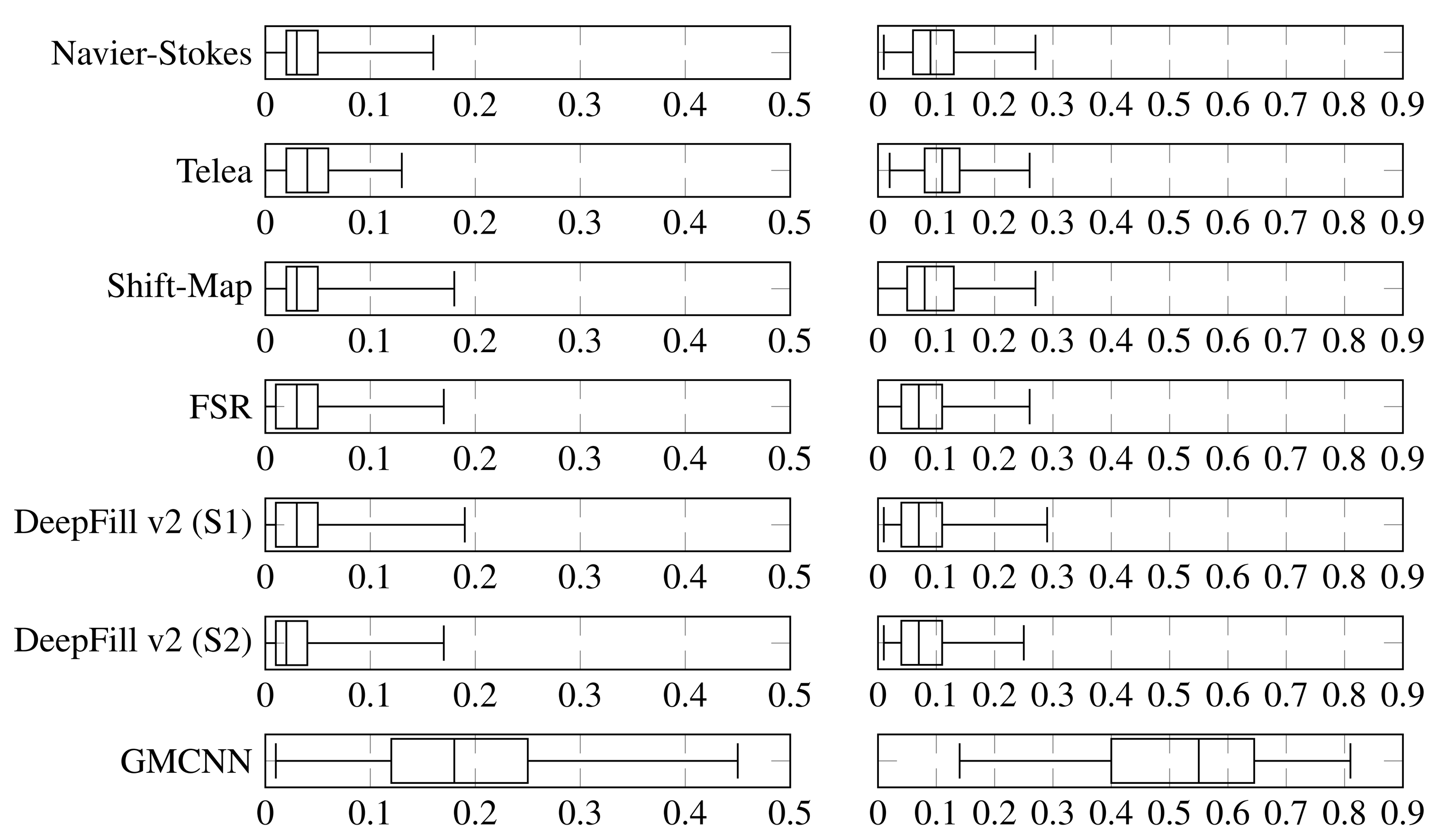
GMCNN Algorithm

The "Generative Multi-column Convolutional Neural Network" [1] (GMCNN) consists of three sub-networks, where only the first one is used for inpainting. The second sub-network implements the local and global discriminators for adversarial training and the third sub-network is a pre-trained VGG network that provides feature values for a feature-based loss.

GROUND TRUTH AND TRAINING DATA

We tested all algorithms on facade images from a textured 3D model of the city of Krefeld in Germany. The ground truth consisted of 206 images that were free of occlusions. These images were assigned various occlusion masks showing trees and different geometric shapes in various sizes.

The neural networks were pre-trained with images of the "places2' dataset. Then transfer learning was applied based on the "Ecole Centrale Paris Facades Database", "FaSyn13", and "CMP" datasets.



(a) distances measured on the entire image

(b) distances measured on occluded regions

Figure 1: Distribution of normalized l_2 distances between ground truth and inpainted images of the entire test dataset; DeepFill v2 (S1) relates to equally weighted loss components and DeepFill v2 (S2) shows the result for a higher weighted l_1 loss

RESULTS

DeepFill v2 delivered excellent results, but also the Shift-Map algorithm performed well. This is expected to be true for the FSR algorithm as well, once a stable implementation is available. The algorithms can be used without problem specific adjustments. There seems to be no longer a need for highly specialized facade inpainting algorithms.

Image	Navier-Stokes	Telea	Shift-Map	FSR	DeepFill v2 (S1)	DeepFill v2 (S2)	GMCNN
1	0.06	0.117	0.051	**0.014**	0.029	0.027	0.589
2	0.059	0.061	0.086	**0.047**	0.054	0.059	0.696
3	0.126	0.133	0.215	0.097	0.094	**0.082**	0.283
4	0.09	0.107	0.089	0.055	0.053	**0.049**	0.574
5	0.168	0.157	0.13	**0.108**	0.174	0.146	0.463
6	0.194	0.175	0.263	0.228	0.168	**0.157**	0.42
7	0.123	0.143	0.155	**0.073**	0.09	0.086	0.492
8	0.115	0.119	0.107	**0.073**	0.102	0.101	0.542
9	0.18	0.17	**0.145**	0.178	0.17	0.166	0.5
10	0.219	0.213	0.222	**0.19**	0.211	0.198	0.537
11	0.175	0.164	0.2	**0.136**	0.185	0.149	0.485
12	0.136	0.128	**0.069**	0.13	0.098	0.089	0.365
13	0.016	0.026	0.013	**0.006**	0.017	0.018	0.795

Table 1: Normalized l_2-distance between ground truth images and inpainted images restricted to the occluded regions. Bold and underlined numbers indicate best and second best results. Numbers relate to images in Figure 2

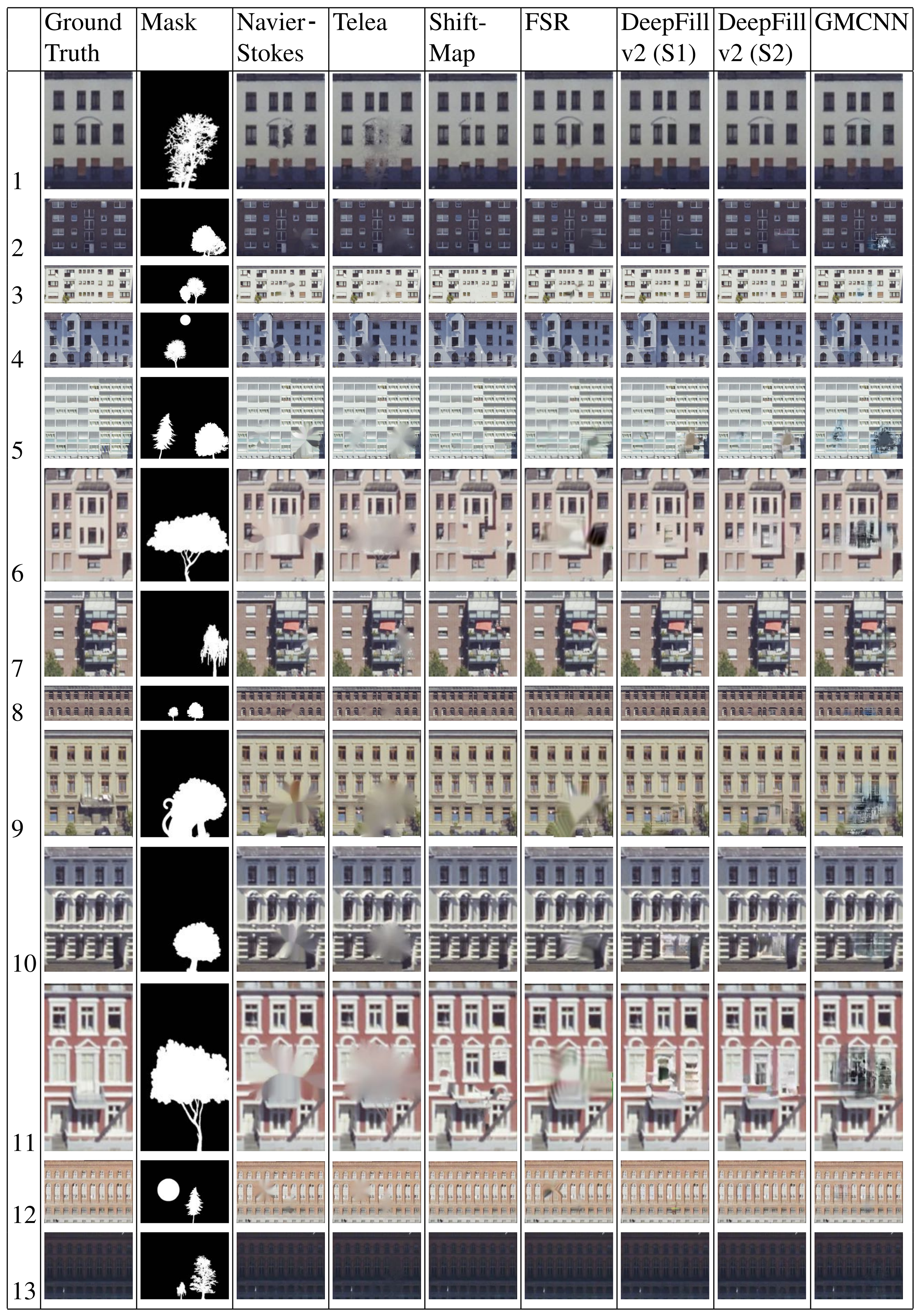


Figure 2: Comparison of inpainting algorithms

[1] Y. Wang et al. "Image Inpainting via Generative Multi-column Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. 2018, pp. 331–340.

[2] J. Yu et al. "Free-Form Image Inpainting with Gated Convolution". In: *Proc. CVPR 2019, arXiv preprint 1806.03589*. 2019.