Image Inpainting

Comparison and Evaluation of Different Architectures

Federico Chiarello - Sara Nanni

The pipeline adopted has allowed to initially consider naive approaches, regarded as baseline, and then integrate increasingly complex models, until analyzing the most emerging ones

Pipeline

incremental methodology

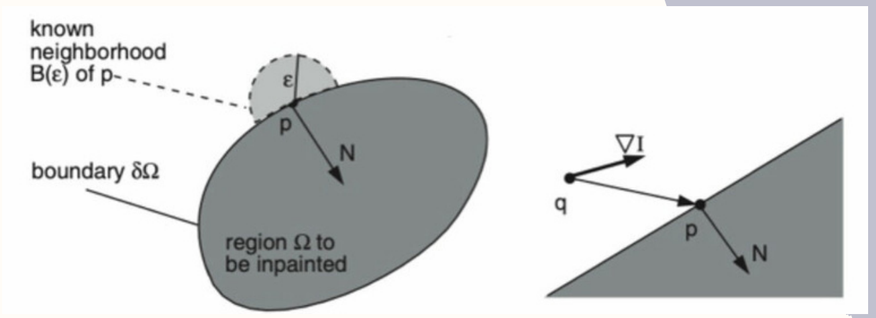
AIM: analyze interesting inpainting approaches, implemented through different model architectures

How: Goal is to investigate composition of traditional computer vision structures with new powerful element.

1. Models Selection
2. Dataset Choice
3. Mask Generation
4. Metric Identification
5. Testing Phase
6. Quantitative and Qualitative Analysis

Baseline

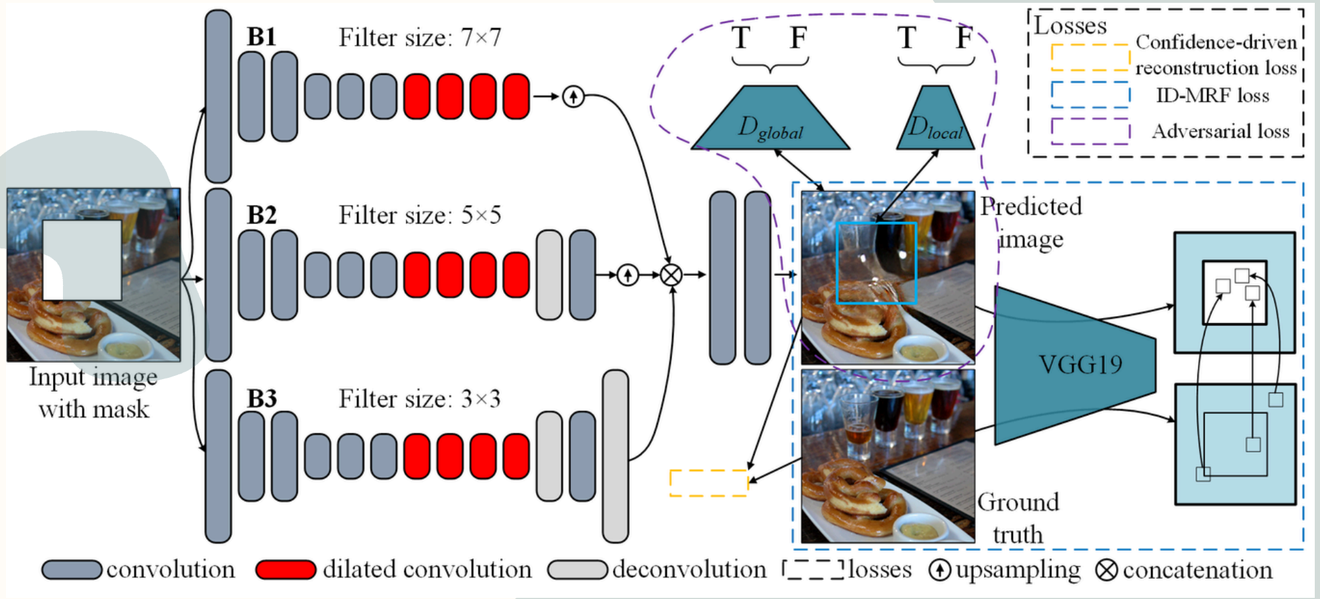
- implementation of the **Telea** inpainting function, from OpenCV library
- smoothness estimator along the image gradient
- weighted average over a known image neighborhood of the corrupted region
- use of Fast Marching Method (**FMM**) to maintain the narrow band that separates the known from the unknown image

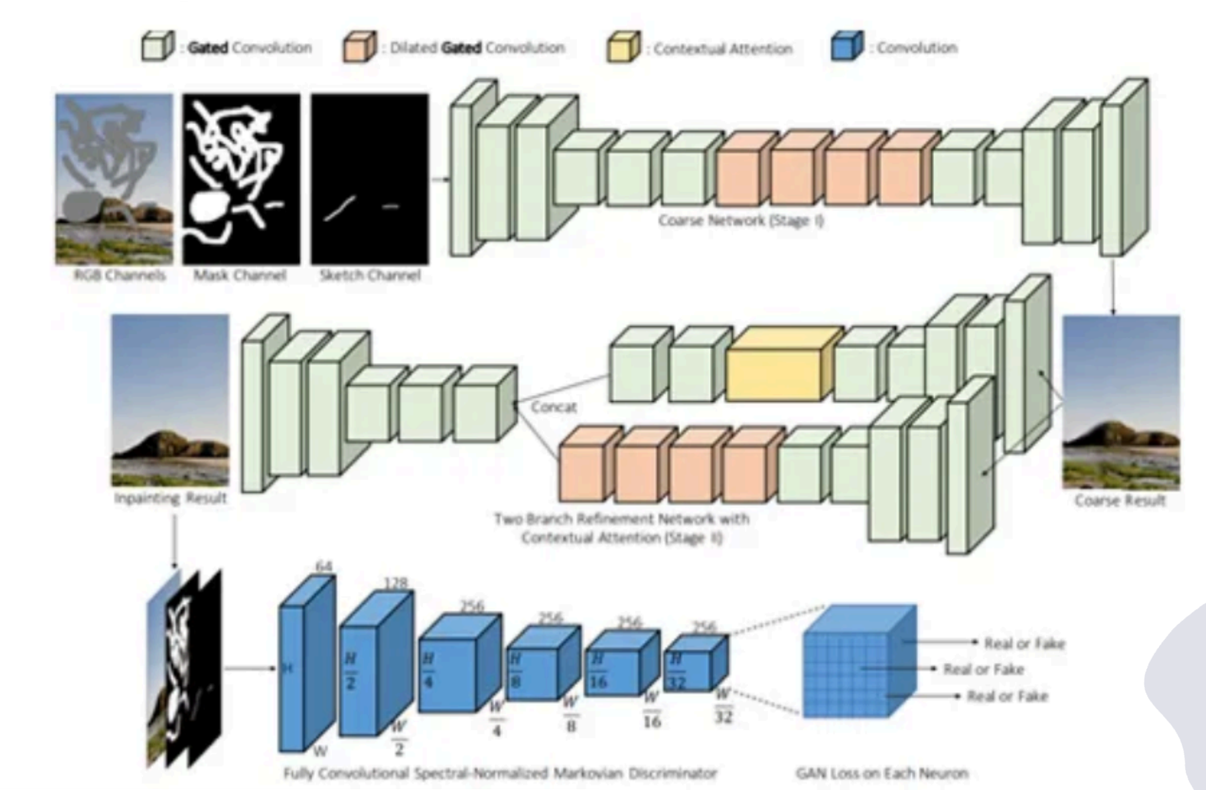


GMCNN

Consists of 3 subnetworks:

- **Generator:** consisting in 3 parallel encoder-decoder branches
- **Global & Local discriminators** for adversarial training
- **Pretraind VGG network** to compute ID-MRF loss



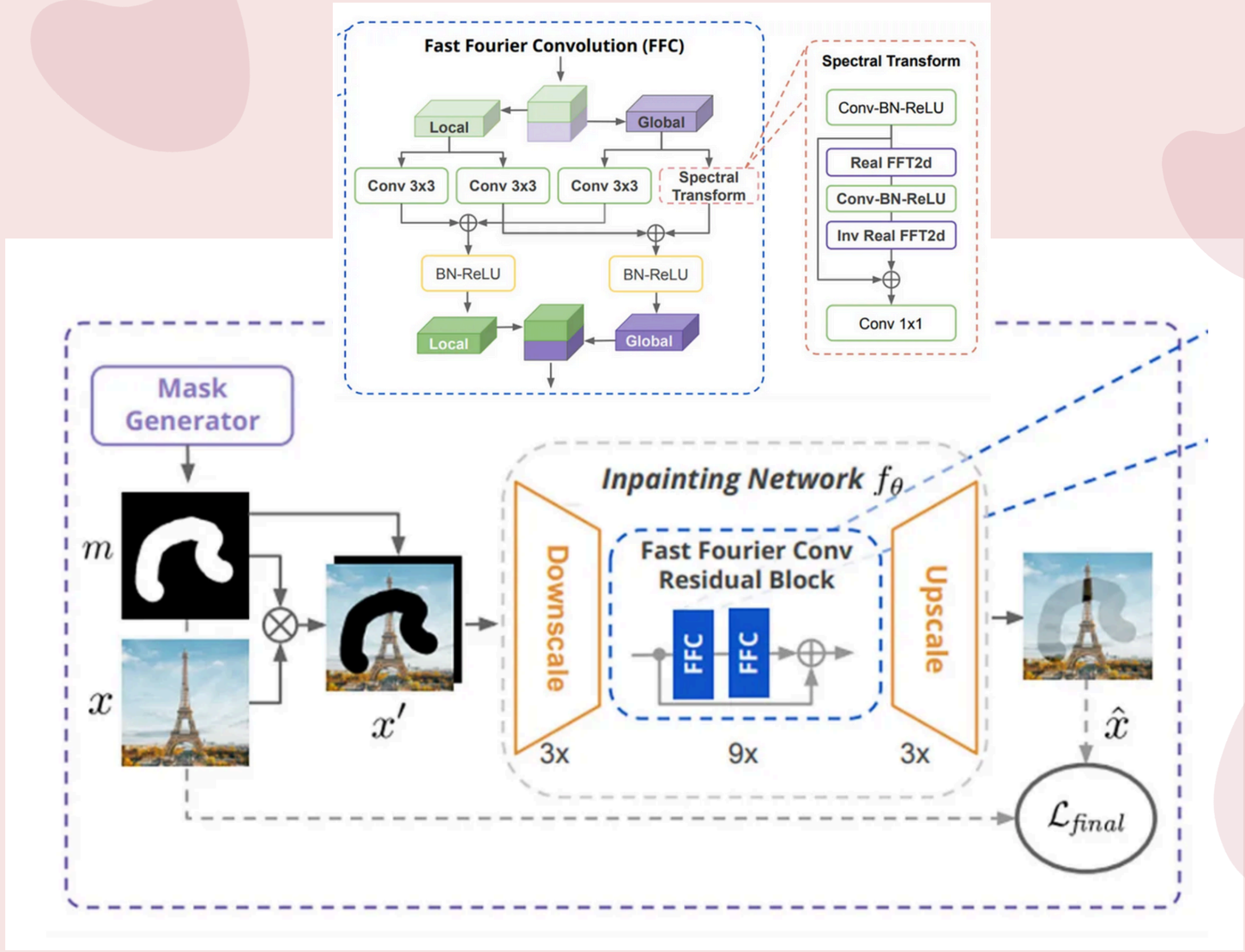


DeepFill v2

- **Contextual Attention**: allows the generator to reconstruct the missing pixels and includes SN-PatchGAN, by applying spectral-normalized discriminator on dense image patches

- **Gated Convolution**: dynamic feature selection mechanism for each channel at each spatial location across all layers

LaMa



Fast Fourier Convolutions

new operator that allows to have a receptive field which covers the entire image, based on a channel-wise fast Fourier Transform

Perceptual Loss Function

ensures that generated areas fit into the global structure of the overall image and that details fills correctly at a local level. Implemented as a weighted sum of several losses: adversarial loss, high receptive field perceptual loss, discriminator-based perceptual loss and R1 gradient penalty

Large Masks

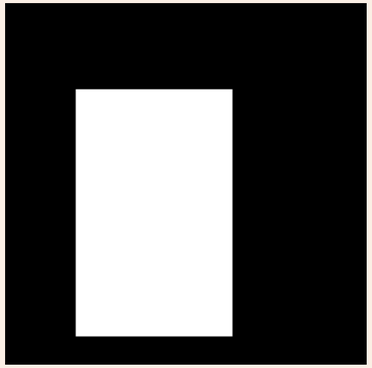
force the network to unlock the potential of the first two components

Places2

- collection of more than 10 million place images from over 400 unique scene categories
- already splitted into training, validation and test sets
- *Places365-Standard* includes 1.8 million train images from 365 scene categories and for each one there are 50 images in the validation set and 900 images in the testing set
- *test-256*: The images in the this archives have already been resized to 256x256 regardless of the original aspect ratio

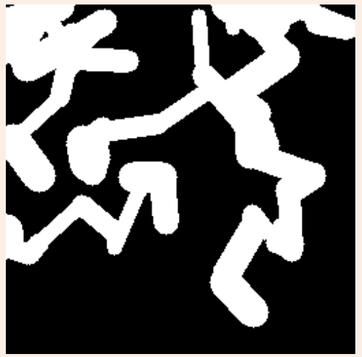
Masks

Seven sizes are available for each mask type:
5, 10, 20, 30, 40, 50 and 60 of covered-image percentage. The choice of testing different mask sizes allows to perform a stress-test on the models capabilities.



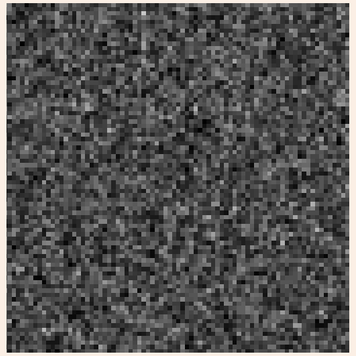
Rectangular

rectangular shaped mask placed randomly into the original image surface



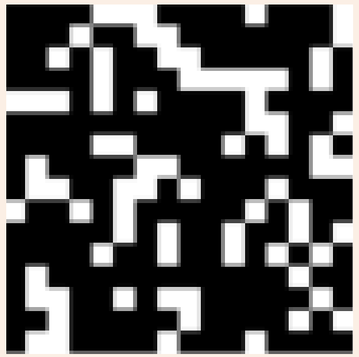
Stroke

free shape mask that simulates a random scribble on the image surface



Random Noise

built randomly masking image pixels resembling the so called salt-and-pepper-noise

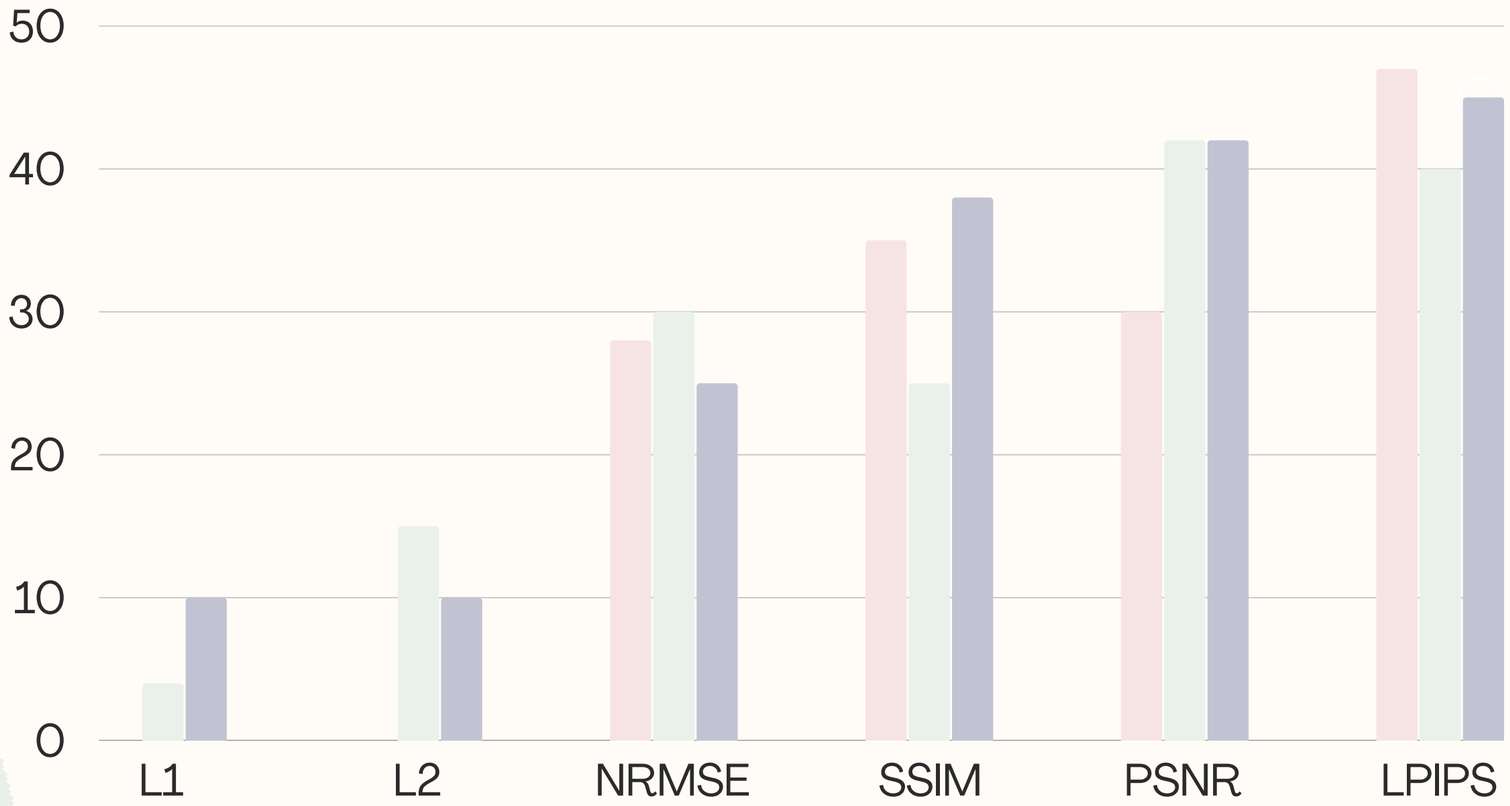


Mosaic

mosaic mask pattern that covers a specified percentage of the image

Metrics

pixel-wise accuracy is not a good metric because it focuses on exact matches between corresponding pixels of the orignal and inpainted images, without account for human perception, more sensitive to overall structures, textures, and coherence in a picture



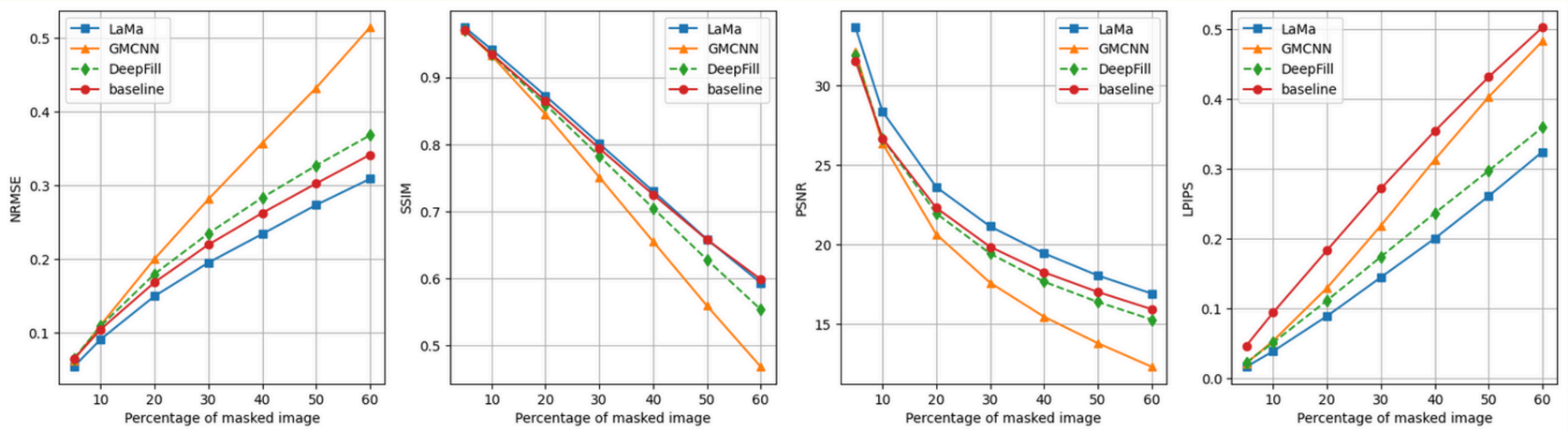
NRMSE: metric based on the RMSE distance, which gives a sense of the average magnitude of the error in pixel values

SSIM: focuses on comparing the structure, luminance, and contrast of the images, which aligns more closely with human visual perception

PSNR: difference between two images by comparing the maximum possible signal power, to the power of the noise

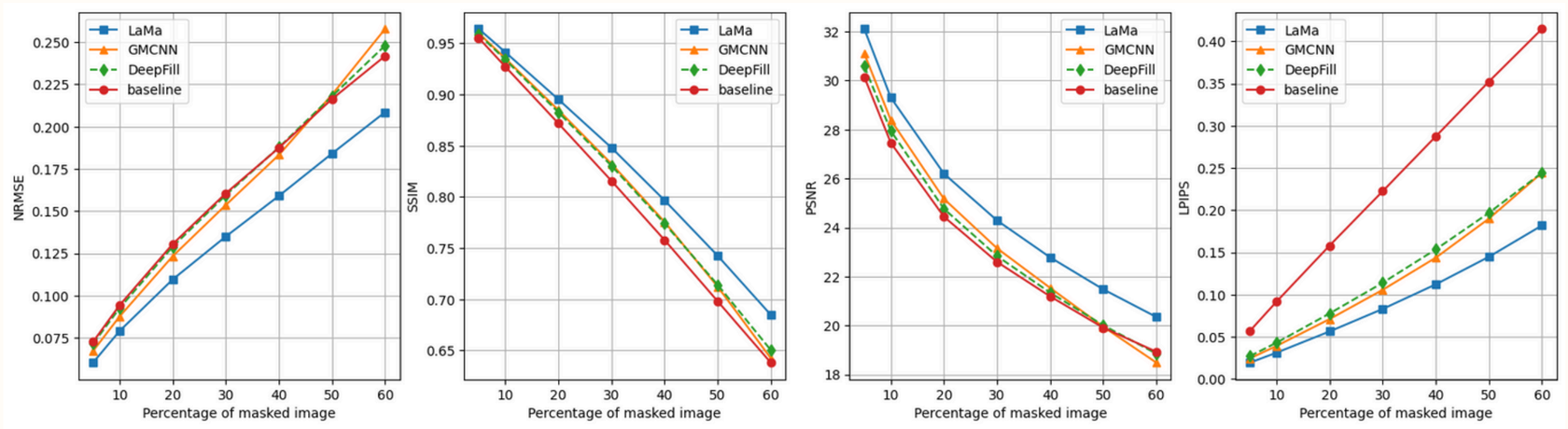
LPIPS: metric based on deep learning models, AlexNet, to capture perceptual differences between images in a way that aligns more closely with human visual perception

Rectangular Masks Metrics Results



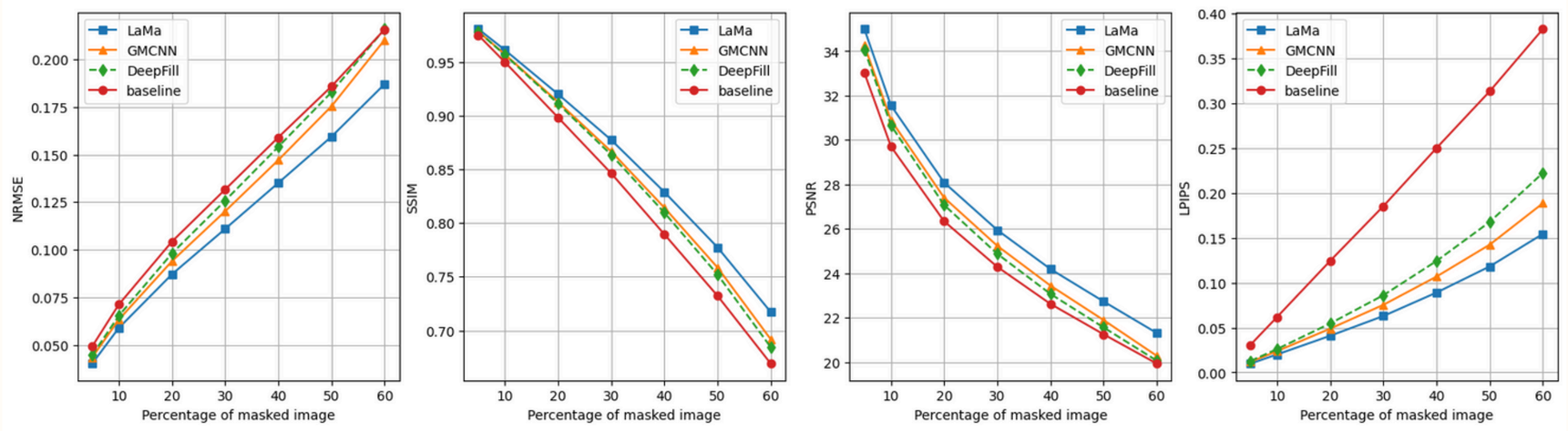
LaMa emerges as the best model, followed by DeepFill, which in turn is surprisingly followed by the baseline, while the GMCNN model seems to be the worst one when dealing with large missing areas.

Stroke Masks Metrics Results



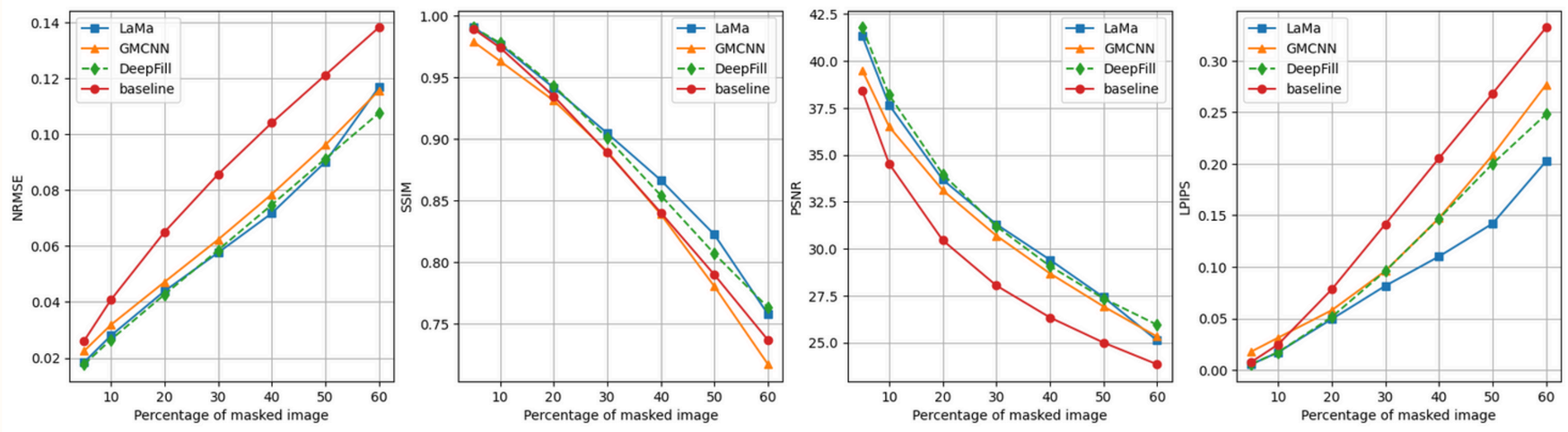
Lama confirms to be the best model, DeepFill is joined by GM-CNN and most of the times overtaken by it, while the baseline always remains below all the other curves

Mosaic Masks Metrics Results



Similar behaviour to Stroke Masks

Random Masks Metrics Results



it’s possible to notice a strange behaviour: Lama and DeepFill curves proceed side by side, GMCNN is positioned on average just below the previous models, with the exception of the SSIM where it is surpassed by the baseline

Test Results Table

metrics scores are shown for the rectangular and stroke masks scenarios, considering three different masks sizes: 10%, 30%, 50%

Rectangular Masks												
	LPIPS			**NRMSE**			**SSIM**			**PSNR**		
	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%
LaMa	0.038	0.144	0.260	0.091	0.195	0.273	0.941	0.801	0.658	28.38	21.13	18.04
DeepFill	0.050	0.173	0.296	0.109	0.234	0.327	0.934	0.782	0.628	26.69	19.42	16.37
GMCNN	0.053	0.218	0.402	0.110	0.281	0.432	0.932	0.751	0.559	26.37	17.58	13.77
Baseline	0.093	0.271	0.430	0.104	0.219	0.302	0.934	0.793	0.657	26.68	19.84	17.00
Stroke Masks												
	LPIPS			**NRMSE**			**SSIM**			**PSNR**		
	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%
LaMa	0.031	0.083	0.144	0.079	0.134	0.184	0.941	0.848	0.742	29.30	24.29	21.47
DeepFill	0.043	0.114	0.196	0.092	0.159	0.217	0.934	0.830	0.714	27.94	22.83	20.00
GMCNN	0.039	0.105	0.190	0.087	0.153	0.218	0.935	0.832	0.712	28.39	23.15	19.95
Baseline	0.091	0.222	0.352	0.094	0.160	0.216	0.927	0.815	0.698	27.45	22.59	19.91

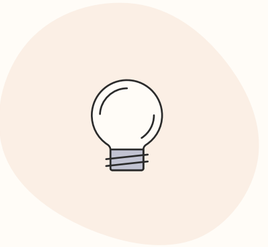

Original
Mask
LaMa
DeepFill
GMCNN
baseline


Original
Mask
LaMa
DeepFill
GMCNN
baseline

Conclusion

Best metric: LPIPS

Since it recognizes the ability of models to faithfully reconstruct details or to insert new textures to uniform the background, it allows to distinguish them from the banal approach adopted by the baseline, which shows a notable gap from the other curves



Best model: LaMa

Overall LaMa proves to be the most popular and most broadly used approach for image inpainting tasks, thanks to Fast Fourier Convolutions (FFCs), which allow for a larger effective receptive field that covers an entire image even in the early layers of the network

Thank you!