Image Inpainting Challenge

NTIRE Workshop and Challenges @ CVPR 2022

Andrés Romero[1], Angela Castillo[2], Jose Abril-Nova[2], Radu Timofte[1,3]

[1] Computer Vision Lab, ETH Zürich

[2] Center for Research and Formation in Artificial Intelligence, Uniandes – Colombia

[3] University of Würzburg, Germany

Agenda

1. Motivation
2. Challenge
    1. Overview
    2. Masks
    3. Datasets
    4. Evaluation
    5. Challenge phases
3. Challenge Methods

MOTIVATION

- Generative models have been successfully used in image inversion problems such as **super-resolution** and **image restoration**. However, unlike these tasks, image inpainting **lacks a standardized** benchmark and evaluation.

- GAN-based inpainting solutions exhibit outstanding results in **object removal** or **texture synthesis**. Nevertheless, these methods struggle with **hallucinating new faces** or producing **semantically coherent** objects.

- The contribution of our challenge to the community is twofold.
  - **Standardize** a set of different and challenging masks, including strokes, image completion, and image extrapolation.
  - **Include** a benchmark of different **scene representations** such as faces, objects, landscapes, and creative art.

THE CHALLENGE
Overview

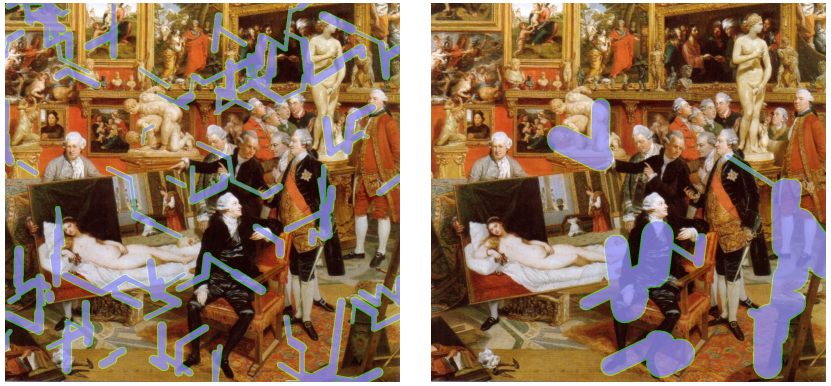
- The goals of the challenge are:
  - Direct and easy **comparison** against recent state-of-the-art Image Inpainting solutions.
  - To perform a **comprehensive analysis** on the different types of masks, for instance, strokes, completion, and nearest neighbor upsampling, among others.
  - To set a **public benchmark** on four different datasets: Portraits, Places, ImageNet, and WikiArt.

- Due to the **size** of the occlusion mask and the **lack of prior** information about the scene, it is unrealistic to generate an ideal solution that faithfully resembles the original image.
- We established two tracks.
  - Track 1: the unsupervised image inpainting track, where no conditional information of the scene is used.
  - Track 2: the semantically-guided image inpainting track, where a semantic segmentation mask is used to guide the inpainting solution.

THE CHALLENGE

Masks

- Strokes:

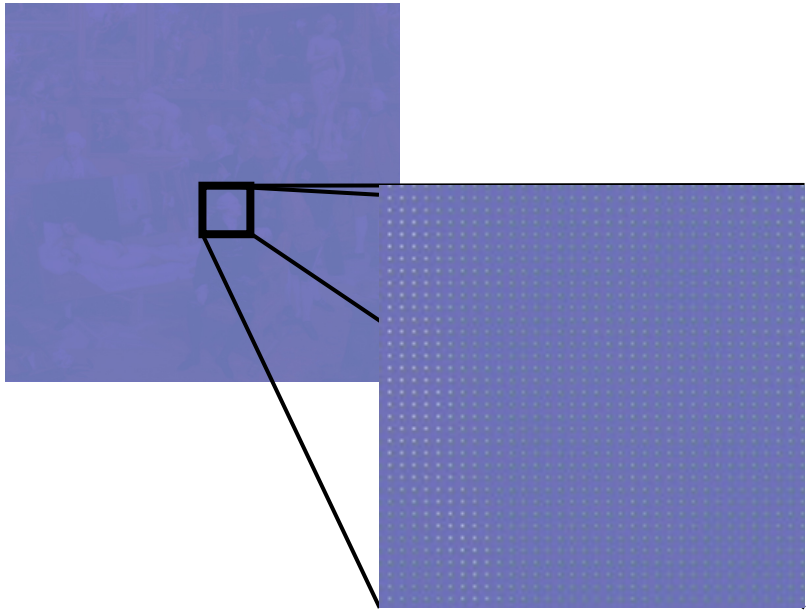


- Image completion:

- Every N Lines:

- Image Expansion:

- Nearest Neighbor:

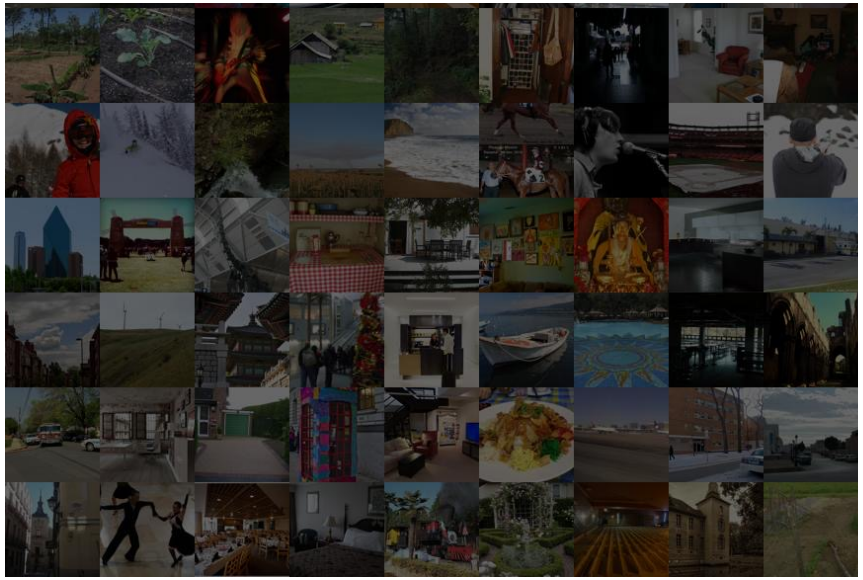


THE CHALLENGE

Datasets

FFHQ[1]

Image inpainting over portraits is a popular application of image inpainting due to the impact on image editings, such as hair replacement, eyeglasses imposition, artifact removal, and smile adjusting.



Places[2]

This dataset was created for deep scene understanding, which collects a categorical dataset with highly diverse and complex scenes such as indoor, nature, urban, street, and rainforest.

1. Tero Karras, Samuli Laine, and Timo Aila. A style- based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 4401–4410, 2019.
2. Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6):1452–1464, 2017.

THE CHALLENGE

Datasets

ImageNet[1]

We employ the ImageNet dataset to analyze the inpainting on more structured and semantic objects.

WikiArt[2]

Our hypothesis is that hallucinating an essential region of a painting requires a deeper understanding of the artist's technique, the context, and the painter's intention.

1. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.
2. K. Nichol. Painter by numbers, wikiart. https://www. kaggle . com / competitions / painter - by - numbers/, 2016.

THE CHALLENGE

Evaluation

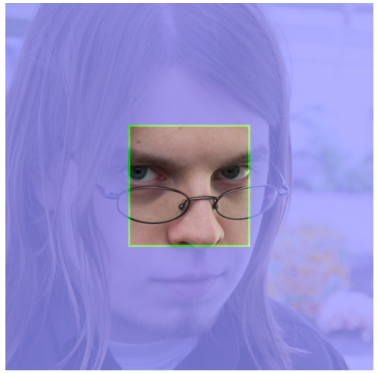
Track 1:

Input

Input

Track 2:

Input



Segmentation map

We compute the mean Intersection over Union (mIoU) with reference to the GT semantic labels.

Metrics

<u>Perceptual:</u>

Learned Perceptual Image Patch Similarity (LPIPS) [1]
Frechet Inception Distance (FID) [2]

<u>Fidelity:</u>

Peak Signal to Noise Ratio (PSNR)
Structural Similarity (SSIM)

As a final ranking, we will select the champion based on the perceptual metrics and a Mean Opinion Score (MOS) for the top solutions.

1. Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. CVPR, 2018.
2. Martin Heusel, Hubert Ramsauer, Thomas Un- terthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017.

THE CHALLENGE

Challenge Phases

Development

The participants get access to the data

Validation

The participants can upload their solutions to the remote server to check the fidelity scores on the validation dataset

Testing

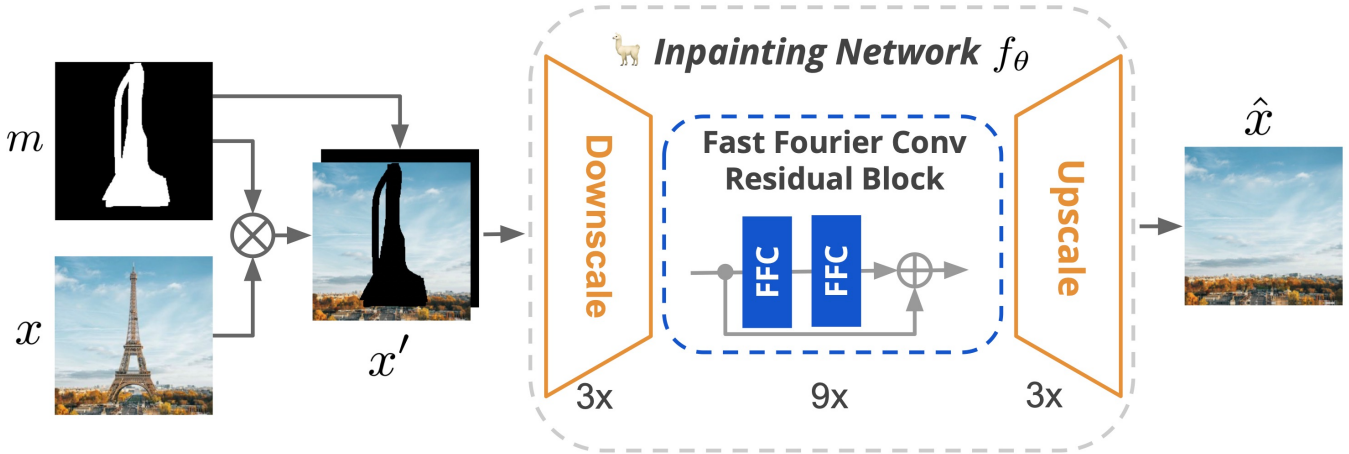
The participants submit their final results, codes, and factsheets

The participants did not have access to the test dataset to avoid issues related to model overfitting, reproducibility of the results, and consistency of the obtained performance values.

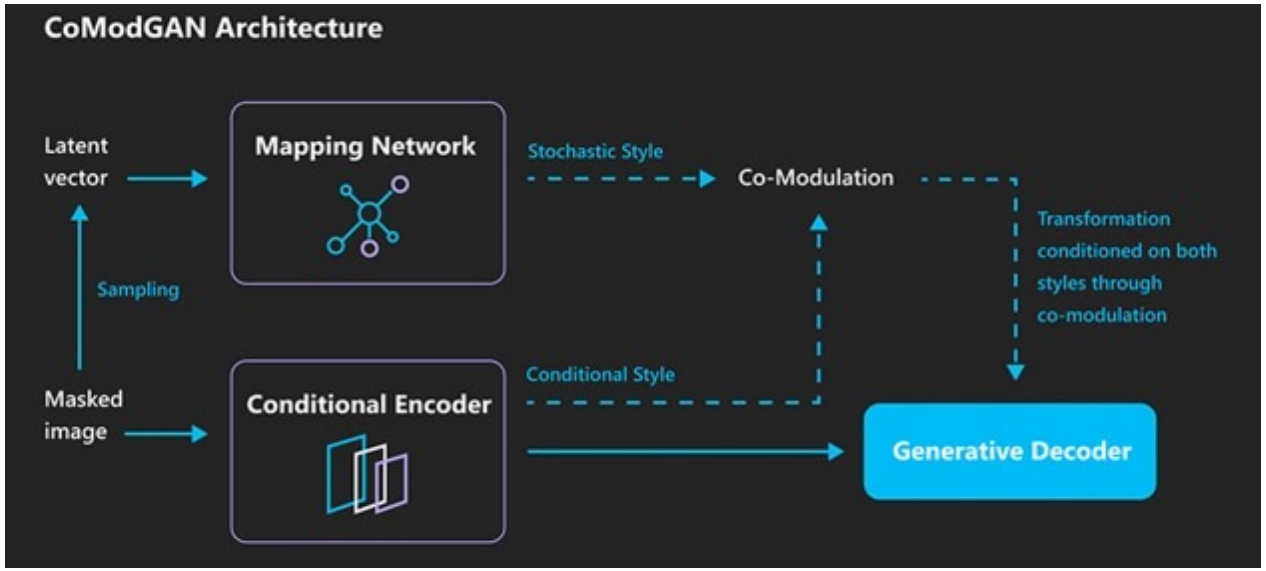
CHALLENGE RESULTS

Baselines

LaMa[1]



CoModGAN[2]



1. Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. WACV22
2. Shengyu Zhao, Jonathan Cui, Yilun Sheng, Yue Dong, Xiao Liang, Eric I Chang, and Yan Xu. Large scale im- age completion via co-modulated generative adversarial networks. arXiv preprint arXiv:2103.10428, 2021.

CHALLENGE RESULTS

Challenge Methods

Track 1

- On this track, all methods were GAN-based. The top 3 solutions used CoModGAN and LaMa.

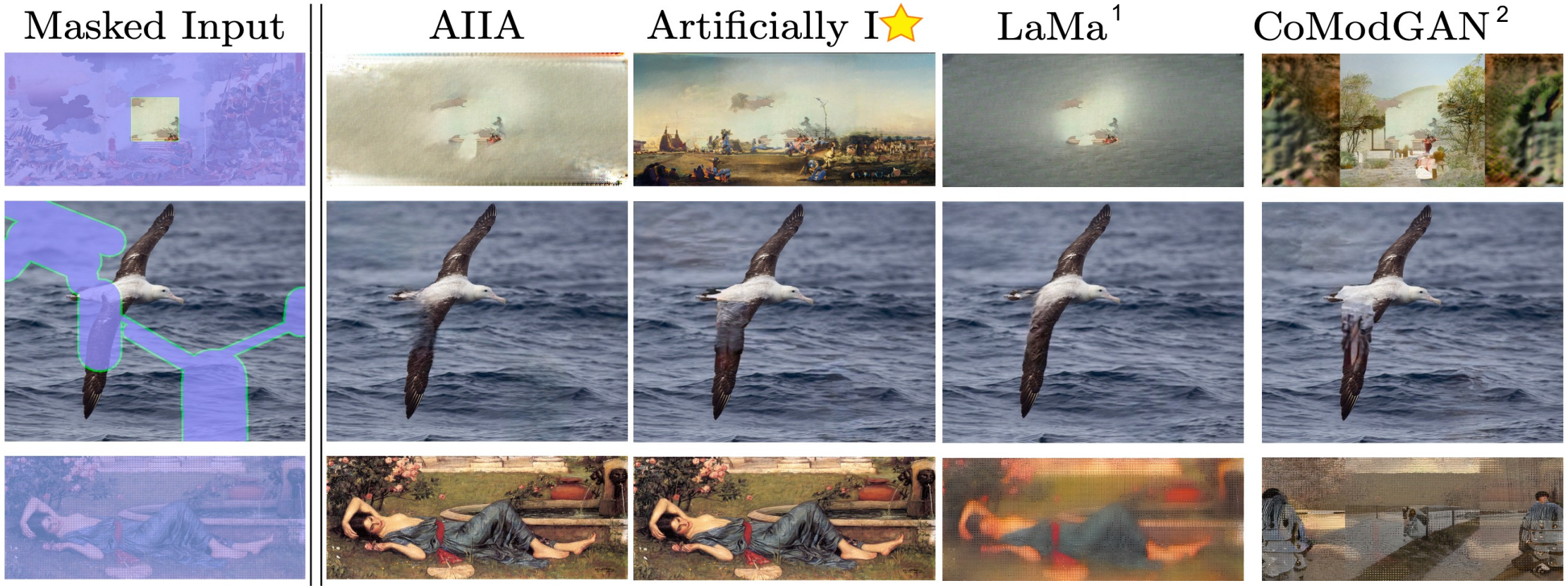
- The winner team, named Artificially Inspired, was based on CoModGAN. This method used more general masks using the pre-trained models as a baseline.

- The runner-up solution, AIIA, exhibited surprisingly good results in some instances surpassing the winner method. AIIA achieved better perceptual in some masks.

CHALLENGE RESULTS

Challenge Methods

Track 1



1. Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. WACV22
2. Shengyu Zhao, Jonathan Cui, Yilun Sheng, Yue Dong, Xiao Liang, Eric I Chang, and Yan Xu. Large scale im- age completion via co-modulated generative adversarial networks. arXiv preprint arXiv:2103.10428, 2021.

CHALLENGE RESULTS

Challenge Methods

Track 1

	Team	Author	LPIPS↓	PSNR↑
FFHQ	AIIA	Zeyu Lu	0.172 ± 0.173	25.316 ± 8.307
	HSSLAB	Rengang Li	0.171 ± 0.185	25.187 ± 8.864
	KwaiInpainting	Jiayin Cai	0.213 ± 0.204	25.060 ± 8.669
	ArtificiallyInspired⭐	Ritwik Das	0.164 ± 0.181	25.999 ± 10.597
	SIGMA	Xiaoqiang Zhou	0.178 ± 0.161	24.860 ± 8.064
	CoModGan		0.546 ± 0.230	10.652 ± 3.815
	LaMa		0.484 ± 0.229	11.152 ± 4.325
Places	AIIA	Zeyu Lu	0.193 ± 0.209	24.145 ± 8.307
	HSSLAB	Rengang Li	0.191 ± 0.217	24.345 ± 8.273
	KwaiInpainting	Jiayin Cai	0.239 ± 0.193	23.410 ± 7.892
	ArtificiallyInspired⭐	Ritwik Das	0.204 ± 0.207	23.248 ± 9.477
	SIGMA	Xiaoqiang Zhou	0.223 ± 0.189	22.562 ± 7.162
	CoModGan		0.496 ± 0.244	11.403 ± 4.154
	LaMa		0.523 ± 0.236	11.184 ± 3.758

CHALLENGE RESULTS

Challenge Methods

Track 2

- The final decision to select the winner was more difficult in this track since the top methods showed good performance.

- The top 2 teams proposed different paradigms to include the semantic information; however, both methods ensure a more faithful semantic reconstruction. These solutions were built upon Diffusion Models and GANs.

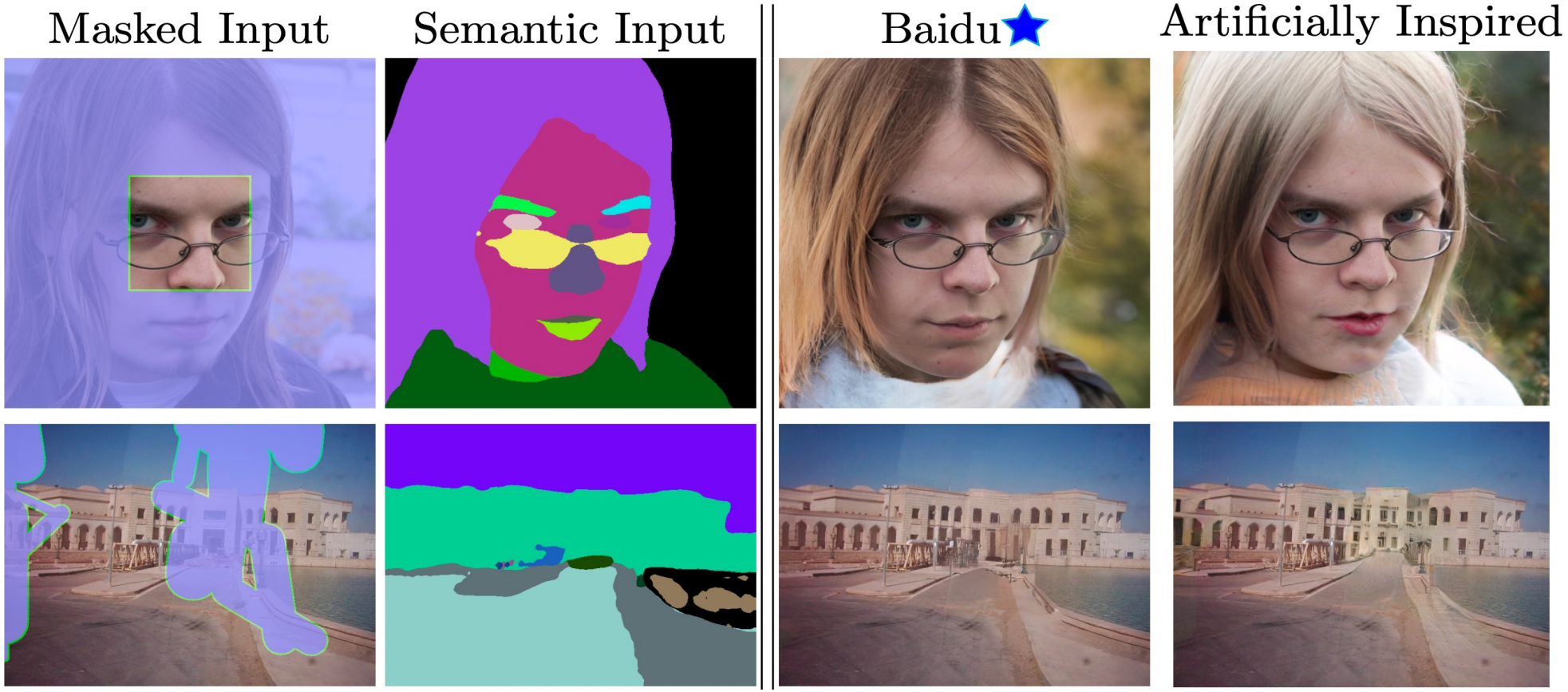
- The winner method, Baidu, leverages the capacity of Latent Diffusion Models to process on the latent representation rather than on the pixel level.

- Artificially Inspired, the second-best method, encoded the semantic information as a styled, while Baidu enforced the network to represent the semantic information explicitly.

CHALLENGE RESULTS

Challenge Methods

Track 2



CHALLENGE RESULTS

Challenge Methods

Track 2

	Team	Author	LPIPS↓	PSNR↑	mIoU↑	MOS↑
FFHQ	MGTV	Xinying Wang	0.134 ± 0.131	25.769 ± 8.817	0.962	4.267
	Baidu★	Zhihong Pan	0.123 ± 0.132	26.254 ± 8.892	0.962	4.582
	HSSLAB	Rengang Li	0.171 ± 0.185	25.187 ± 8.864	0.826	-
	ArtificiallyInspired	Ritwik Das	0.138 ± 0.136	26.827 ± 9.864	0.948	4.575
Places	MGTV	Xinying Wang	0.193 ± 0.180	23.738 ± 7.838	0.672	3.546
	Baidu★	Zhihong Pan	0.182 ± 0.188	23.289 ± 8.293	0.636	3.882
	HSSLAB	Rengang Li	0.191 ± 0.217	24.345 ± 8.273	0.574	-
	ArtificiallyInspired	Ritwik Das	0.188 ± 0.174	23.868 ± 8.864	0.655	3.700

FINAL REMARKS

- For track 1, we notice the influence of the masks in the generation process. If the masks cover meaningful semantic regions (e.g., the eyes), the models tend to lower the perceptual performance because it is harder to hallucinate these structures.

- On track 2, the general performance improves due to the semantic guidance given by the maps. However, this track is more competitive because faithful reconstruction acquires more relevance.

- This challenge establishes a benchmark to push the study of image inpainting. We hope this work will robustify image inpainting methods while helping us comprehend new open challenges in this area.

Acknowledgments

We thank the NTIRE 2022 sponsors: Huawei, Reality Labs, Bending Spoons, MediaTek, OPPO, Oddity, Voyage81, ETH Zürich (Computer Vision Lab) and University of Würzburg (CAIDAS).