NTIRE 2022 Image Inpainting Challenge: Report

Andrés Romero Angela Castillo Jose Abril-Nova Radu Timofte Ritwik Das
Sanchit Hira Zhihong Pan Min Zhang Baopu Li Dongliang He Tianwei Lin
Fu Li Chengyue Wu Xianming Liu Xinying Wang Yi Yu Jie Yang
Rengang Li Yaqian Zhao Zhenhua Guo Baoyu Fan Xiaochuan Li
Runze Zhang Zeyu Lu Junqin Huang Gang Wu Junjun Jiang Jiayin Cai
Changlin Li Xin Tao Yu-Wing Tai Xiaoqiang Zhou Huaibo Huang

Abstract

Image Inpainting has recently become an important research problem due to the rise of generative image synthesis models. While many solutions have been proposed for this problem, it is challenging to establish a testbed due to the different possible types of inpainting masks e.g., *completion mask, expand mask, thick brushes mask,* etc. *Most inpainting solutions shine on object removal or texture synthesis, while semantic generation is still difficult to achieve. To address these issues, we introduce the first general Image Inpainting Challenge. The target is to develop solutions that can achieve a robust performance across different and challenging masks while generating compelling semantic images. The proposed challenge consists of two tracks: unsupervised image inpainting and semantically-guided image inpainting. For Track 1, the participants were provided with four datasets: FFHQ, Places, ImageNet, and WikiArt, and trained their models to perform a mask-agnostic image inpainting solution. For Track 2, FFHQ and Places only. This report gathers the description and discussion of all solutions that participated in the final stage of the challenge.*

1. Introduction

Image inpainting is the task of filling in missing information in an image. Usually, these regions are the outcome of human intervention, such as object removal or image editing, as well as degradation or artifacts. The main goal of image inpainting is to produce realistic and pleasant images that harmonize well with the rest of the image.

In recent years there has been a substantial increase in published papers using generative models. Since the introduction of GANs [13], several solutions for image inversion problems such as super-resolutions [56, 67], image restoration [65, 24], and image inpainting [60, 50] have been successfully applied with impressive results. However, unlike super-resolution or restoration problems where there are clear benchmarks and evaluation protocols [1, 33], image inpainting merely relies on a binary mask that guides the inpainted region, which hinders a standardization. Therefore, even though state-of-the-art methods directly compare to others on the same dataset, the binary masks often vary from method to method.

Furthermore, there is a trend in GAN-based inpainting solutions due to the fast inference and relatively fast training. Recent GAN-based methods [42, 63, 26, 37, 19, 34, 57, 58, 15, 25, 61, 50, 18, 64] produce outstanding results at object removal or texture synthesis. However, hallucinating new faces, such as in the image completion task, or producing semantically coherent objects, such as the image expansion task, is still very challenging.

Noteworthy, VAE-based methods [70, 68], systems leveraging Deep Generative priors [53, 69, 43] (*e.g.* StyleGAN [21]), Auto-Regressive methods [62, 40, 54], and Diffusion Models [49, 9, 31, 35, 47, 28, 44] have been successfully applied to image inpainting.

This Image Inpainting Challenge contributes with two main advantages for the community. First, standardize a set of different and challenging masks that include the classical strokes and image completion or image extrapolation. And second, to include a benchmark that consists of different scene representations

*Andrés Romero, Angela Castillo, Jose Abril-Nova, and Radu Timofte *(me@afromero.co, a.castillo13@uniandes.edu.co, jm.abril@uniandes.edu.co, radu.timofte@uni-wuerzburg.de)* are the main NTIRE 2022 Image Inpainting challenge organizers. The other authors participated in the challenge.
Appendix A contains the authors' team names and affiliations.

NTIRE 2022 Workshop website:
https://data.vision.ee.ethz.ch/cvl/ntire22/

Figure 1. Image Inpainting Challenge masks. We selected 7 challenging masks for our challenge. Participants should provide a mask-agnostic model able to perform inpainting under any of the depicted cases. Purple color ■ represents the masked region of the image. *Click on each image for full resolution and zoom-in.*

such as faces, objects, landscapes, and creative art.

Jointly with the NTIRE workshop, we propose an NTIRE challenge on Image Inpainting: the task of predicting the values of missing pixels in an image so that the completed result looks realistic and coherent.

This challenge is a part of the *NTIRE 2022 Challenges*: spectral recovery [3], spectral demosaicing [2], perceptual image quality assessment [14], inpainting [45], night photography rendering [12], efficient super-resolution [23], learning the super-resolution space [29], super-resolution and quality enhancement of compressed video [59], high dynamic range [41], stereo super-resolution [55], burst super-resolution [4].
The results obtained in the other competitions and the description of the proposed solutions can be found in the corresponding challenge papers.

2. Challenge

The goals of this challenge are: (i) Direct and easy comparison of recent state-of-the-art Image Inpainting solutions. (ii) To perform a comprehensive analysis on the different types of masks, for instance, strokes, completion, nearest neighbor upsampling, *etc*. Thus, highlighting the pros and cons of each method for each type of mask. (iii) To set a public benchmark on 4 different datasets: Portraits [21], Places [71], ImageNet [11], and WikiArt [36].

See an example of the different inpainting masks in Figure 1.

2.1. Overview

Image inpainting is an ill-posed problem where the main goal is to reconstruct a photo-realistic image from the filled-with-holes image counterpart. From a classical perspective, an ideal solution should faithfully resemble the original image. However, in many cases, it is unfeasible due to the size of the inpainted mask and the lack of prior information about the scene. Motivated by this observation, we created two tracks for this challenge: (1) the unsupervised image inpainting track, where no conditional information of the scene is used, and (2) the semantically-guided image inpainting track, where a semantic segmentation mask is used for guiding the inpainting solution. The information about the challenge was provided on a public GitHub page: https://github.com/affromero/NTIRE22_Inpainting.

2.2. Masks

Most image inpainting methods [50, 64, 61, 69] are directly trained to solve the inverse problem, which requires specifying the inpainting binary mask during training. Indeed, in some cases the selection of the mask is critical for better performance [50, 61], which dramatically hinders the comparison with similar approaches. This limitation poses an important issue because there is no clear notion of *the ideal mask*. In this challenge, we opt to use the common strategies found in Image Inpainting solutions and more complex mask generations. In total, we use seven different types of masks (see Figure 1):

- *Strokes.* 3 different types of strokes (thin, medium, and thick). These stroke generations are based on the recent LaMa [50] inpainting solution.

- *Image completion.* The mask ranges from 20% to 89% of the image and could be horizontal or vertical.

- *Every N Lines.* We also include a degradation that involves pixel removal from an image. The removal is uniformly either every N lines vertical or horizontal, where N is randomly selected from 2, 3, or 4.

- *Image Expansion.* This mask analyses the extrapolation capabilities of each solution by maintaining only the pixels of a square on the center of an image.

- *Nearest Neighbor.* This mask only leaves pixels with a stride of N in height and width dimension, where N is randomly chosen from 2, 3, or 4. This mask can be seen as an indirect application of image inpainting to super-resolution.

2.3. Dataset

For the unsupervised track, we use 4 different datasets that consists of portraits (FFHQ [21]), scenes (Places [71]), complex objects (ImageNet [11]), and creative art (WikiArt [36]). For the semantically-guided track, we only use portraits and scenes.

FFHQ [21] Image inpainting over portraits is one of the most popular applications of image inpainting due to the impact on image editings, such as hair replacement, eyeglasses imposition, artifact removal, and smile adjusting. The FFHQ consists of 1,028×1,028 high-resolution 70,000 portraits with a high variation in ethnicity, age, and image background. We divide the training, validation, and test using the standard setup: 60,000 for training, 10,000 for validation, and the remaining 10,000 are part of our test challenge set. For Track 2, we use an automated semantic segmentation model [38] to parse the face and use it as conditional information.

Places [71] The Places 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. The entire dataset consists of more than one million images. Recently, many inpainting solutions have used the Places dataset benchmark due to the pleasant and impressive inpainted results in generic object removal. We use the publicly available training and validation set as part of the challenge's training, validation, and test sets. For the Track 2, we use an automated semantic segmentation model, namely DeepLabV3 [7] trained on CocoStuff-164k [6] - MM-segmentation Framework [10], to parse the scene, and use it as conditional information.

ImageNet [11] Following a recent trend in image inpainting solutions, we also employ the ImageNet dataset to analyze the inpainting on more structured and semantic objects.

WikiArt [36] In contrast to the aforementioned datasets, WikiArt has not been a usual benchmark for the image inpainting problem. Therefore, in this challenge, we propose to include a creative art dataset. Our rationale 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. We use the publicly available training and validation set as part of the challenge's training, validation, and test sets.

2.4. Evaluation

Given that Image Inpainting is an inverse problem, in which most solutions differ from the ground truth at pixel level, we use Perceptual Metrics to rank the participants.

We use the Learned Perceptual Image Patch Similarity (LPIPS) [66] and Frechet Inception Distance (FID) [17] as perceptual metrics, as well as fidelity metrics such as the standard Peak Signal to Noise Ratio (PSNR) and the Structural Similarity (SSIM) index as often employed in the literature.

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

- In Track 1, the participants should inpaint the input image according to the input mask, and the evaluation is conducted between the inpainted image and the ground-truth image.

- In Track 2, the participants should inpaint the input image according to the input mask and the semantic input map. The evaluation is conducted between the inpainted and ground-truth images and should be consistent with the semantic mask. Thus, a semantic segmentation network generates the semantic labels of the completed images, and we compute the mean Intersection over Union (mIoU) with reference to the ground-truth semantic labels.

2.5. Challenge Phases

The challenge consisted of the following phases:

I. *Development:* the participants get access to the data;

II. *Validation:* the participants can upload their solutions to the remote server to check the fidelity scores on the validation dataset;

III. *Testing:* the participants submit their final results, codes, and factsheets.

During the final challenge phase, the participants did not have access to the test dataset. Instead, they had to submit their final generated images and the trained models that the challenge organizers subsequently used to check both the perceptual and the fidelity results of each submission under identical conditions. This approach solved all the issues related to model overfitting, reproducibility of the results, and consistency of the obtained performance values.

3. Challenge Results

3.1. Track 1: Unsupervised Image Inpainting

From above 100 registered participants, 5 teams entered the final phase and submitted valid results, codes, executables and factsheets. Tables 1 and 2 summarizes the final challenge results and reports FID, LPIPS, PSNR, and SSIM scores for each submitted solution on the final test dataset and on the validation set, while Figure 2 shows the obtained qualitative results. Additionally, Table 3 reports the MOS scores.

The proposed methods are described in section 5.1, and the team members and affiliations are listed in Appendix A.

3.2. Track 2: Image Inpainting guided by pixel-wise semantic labels

From above 100 registered participants, 4 teams entered the final phase and submitted valid results, codes, executables and factsheets. Tables 4 and 5 summarizes the final challenge results and reports FID, LPIPS, PSNR, SSIM, and mIoU scores for each submitted solution on the final test dataset and on the validation set, while Figure 3 shows the obtained qualitative results. Additionally, Table 6 reports the MOS scores.

The proposed methods are described in section 5.2, and the team members and affiliations are listed in Appendix A.

4. Discussion

4.1. Baselines

We selected two GAN-based state-of-the-art solutions as baselines: LaMa [50] and CoModGAN [69]. We use the publicly available pre-trained weights for FFHQ and Places over all our masks and datasets. However, CoModGAN requires a fixed image size. We bypassed this issue by padding all our images to be 512×512 image size. Therefore, larger images were unfolded and padded accordingly. Due to this CoModGAN preprocessing, images might not have harmonic transitions as in Figure 2. Moreover, as LaMa and CoModGAN models were trained on different masks than those used in the challenge, directly applying their models to our masks lead to a poor performance.

4.2. Track 1

On this track, all methods rely on GANs [13]. Most of the reported solutions were based on existing techniques, such as CoModGAN [69], LaMa [50], and DeepFillv2 [60], being the top 3 solutions based on CoModGAN and LaMa. The champion fine-tunes the original version of CoModGAN to work on the more challenging settings proposed in this challenge.

Interestingly, the runner-up *AIIA* solution, which is based on LaMa, can perform surprisingly well in certain cases, and even the LPIPS is overall better than that of *ArtificiallyInspired*. However, LaMa is a method that employs Fast Fourier Convolution (FFC) [8] layers as the main core. FFC works well on repeated patterns or textures, yet it lacks hallucinating new semantic information, as in the first row of Figure 2. The solution presented by *ArtificiallyInspired* is superior because CoModGAN is a method that employs Modulated Convolutions [22], which has been extensively used for image synthesis problems, where the generation of new semantic information is crucial, hence outperforming other competitors on the more challenging masks.

Tables in Appendix C.1 depict the detailed results for each mask. Image inpainting is an even more ill-posed task if the input mask is completely unknown to the model. Therefore, some masks harm even more the performance of the network compared to others. Specifically, the masks that do not interfere with the image's semantic information exhibit better performance than the rest (*i.e.* Every-N-Lines, Nearest-Neighbor, and Thin Strokes). On the contrary, masks in which the general structure is covered and it is perceptually harder to recognize the semantics, result are harder to deal with, thus lowering the performance (*i.e.* Completion, Expand, and Thick Strokes).

See additional qualitative results in Appendix B.1.

4.3. Track 2

Participants explored very different options to include the semantic information as part of the inpainting process on this Track. In contrast to Track 1, the decision for the winner was difficult, even on the subjective Mean Opinion Score (See Table 6). The top 2 methods employ different paradigms for the inpainting problem: Diffusion Models and GANs. The top solution ensures a more faithful reconstruction of the semantic information, *i.e.* see mIOU column on Table 5.

Recently, Diffusion Models have shown outstanding results in image inpainting [32, 49, 9, 31, 35, 47, 28, 44] as they can be applied to this task without direct supervision. However, they are known for their slow inference time, which hinders their practicability to real-world cases, and in the case of this challenge, to be tested on 7,000 images per dataset. Nevertheless, the winner of the challenge is based on a Latent Diffusion Model (LDM) [44] system, which performs the denoising process on a latent representation instead of the pixel level, which dramatically reduces the inference

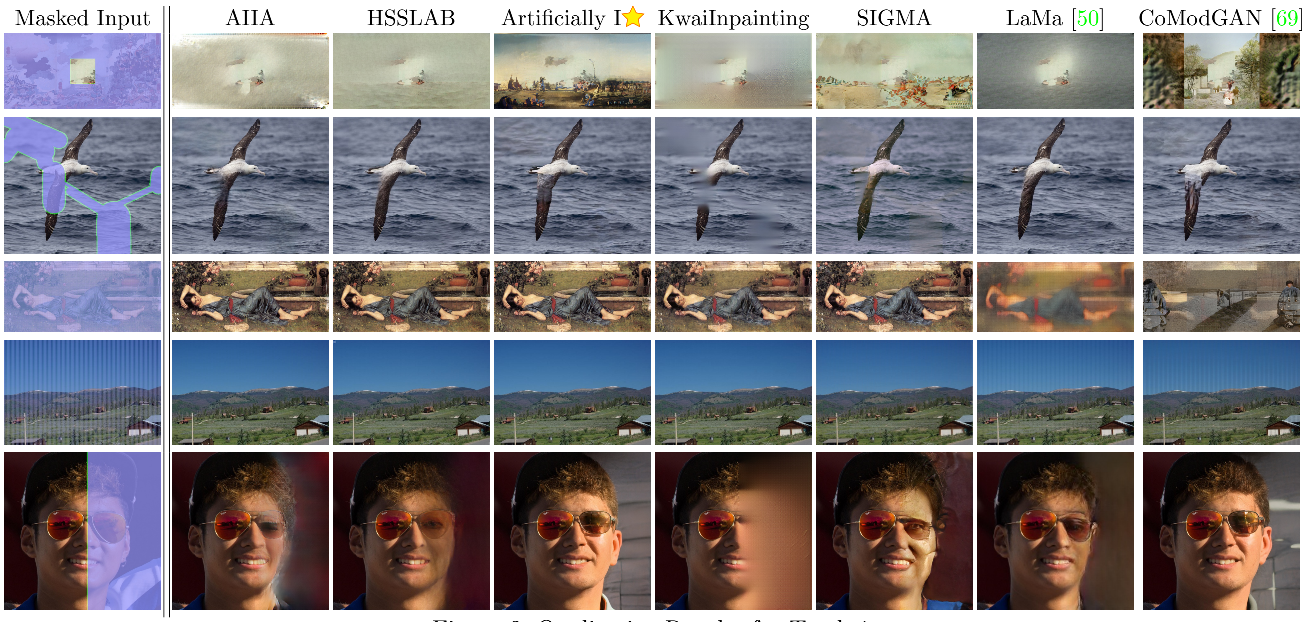

Figure 2. Qualitative Results for Track 1.

	Team	Author	Framework	FID↓	LPIPS↓	PSNR↑	SSIM↑
FFHQ	AIIA	Zeyu Lu	PyTorch	10.107	0.172 ± 0.173	25.284 ± 8.271	0.815 ± 0.154
	HSSLAB	Rengang Li	PyTorch	10.744	0.172 ± 0.186	26.882 ± 9.791	0.852 ± 0.146
	KwaiInpainting	Jiayin Cai	PyTorch	21.290	0.214 ± 0.203	24.995 ± 8.580	0.839 ± 0.146
	ArtificiallyInspired	Ritwik Das	PyTorch & Tensorflow	4.703	0.165 ± 0.181	25.907 ± 10.613	0.816 ± 0.188
	SIGMA	Xiaoqiang Zhou	PyTorch	7.274	0.179 ± 0.161	24.804 ± 8.038	0.795 ± 0.159
	CoModGan [69]			126.037	0.548 ± 0.231	10.553 ± 3.709	0.413 ± 0.307
	LaMa [50]			111.481	0.486 ± 0.229	11.052 ± 4.209	0.446 ± 0.272
Places	AIIA	Zeyu Lu	PyTorch	8.856	0.193 ± 0.209	24.142 ± 8.383	0.800 ± 0.188
	HSSLAB	Rengang Li	PyTorch	9.262	0.190 ± 0.216	25.739 ± 9.249	0.823 ± 0.189
	KwaiInpainting	Jiayin Cai	PyTorch	18.336	0.241 ± 0.217	23.351 ± 8.011	0.787 ± 0.195
	ArtificiallyInspired	Ritwik Das	PyTorch & Tensorflow	7.680	0.205 ± 0.208	23.226 ± 9.595	0.777 ± 0.221
	SIGMA	Xiaoqiang Zhou	PyTorch	11.676	0.224 ± 0.190	22.507 ± 7.211	0.747 ± 0.201
	CoModGan [69]			68.742	0.496 ± 0.243	11.321 ± 4.177	0.477 ± 0.289
	LaMa [50]			68.025	0.525 ± 0.238	11.118 ± 3.861	0.327 ± 0.266
ImageNet	AIIA	Zeyu Lu	PyTorch	10.402	0.181 ± 0.217	25.287 ± 9.197	0.791 ± 0.217
	HSSLAB	Rengang Li	PyTorch	9.010	0.172 ± 0.217	27.239 ± 10.377	0.817 ± 0.216
	KwaiInpainting	Jiayin Cai	PyTorch	18.693	0.237 ± 0.243	23.913 ± 9.002	0.775 ± 0.223
	ArtificiallyInspired	Ritwik Das	PyTorch & Tensorflow	12.546	0.197 ± 0.226	24.343 ± 10.023	0.778 ± 0.238
	SIGMA	Xiaoqiang Zhou	PyTorch	20.251	0.254 ± 0.205	22.401 ± 6.705	0.726 ± 0.229
	CoModGan [69]			40.731	0.454 ± 0.263	12.764 ± 5.507	0.525 ± 0.325
	LaMa [50]			41.721	0.509 ± 0.250	11.586 ± 4.060	0.260 ± 0.231
WikiArt	AIIA	Zeyu Lu	PyTorch	15.031	0.220 ± 0.216	24.204 ± 8.175	0.766 ± 0.202
	HSSLAB	Rengang Li	PyTorch	12.788	0.203 ± 0.211	25.641 ± 8.726	0.798 ± 0.202
	KwaiInpainting	Jiayin Cai	PyTorch	26.498	0.266 ± 0.214	22.925 ± 7.212	0.757 ± 0.204
	ArtificiallyInspired	Ritwik Das	PyTorch & Tensorflow	8.643	0.229 ± 0.215	23.620 ± 9.169	0.757 ± 0.216
	SIGMA	Xiaoqiang Zhou	PyTorch	14.664	0.243 ± 0.191	22.505 ± 6.816	0.718 ± 0.207
	CoModGan [69]			89.204	0.514 ± 0.251	11.136 ± 4.382	0.476 ± 0.291
	LaMa [50]			88.894	0.556 ± 0.242	11.022 ± 4.137	0.292 ± 0.259

Table 1. Quantitative results over the Validation set of Track 1.

time to an average of 10s per 512×512 image size.

In contrast to *Baidu*, *Artificially Inspired* does not enforce the network to represent the semantic information explicitly. It extends the method presented in

	Team	Author	Framework	FID↓	LPIPS↓	PSNR↑	SSIM↑
FFHQ	AIIA	Zeyu Lu	PyTorch	9.823	0.172 ± 0.173	25.316 ± 8.307	0.814 ± 0.155
	HSSLAB	Rengang Li	PyTorch	13.504	0.171 ± 0.185	25.187 ± 8.864	0.821 ± 0.147
	KwaiInpainting	Jiayin Cai	PyTorch	21.345	0.213 ± 0.204	25.060 ± 8.669	0.838 ± 0.147
	ArtificiallyInspired⭐	Ritwik Das	PyTorch & Tensorflow	4.719	0.164 ± 0.181	25.999 ± 10.597	0.816 ± 0.188
	SIGMA	Xiaoqiang Zhou	PyTorch	7.203	0.178 ± 0.161	24.860 ± 8.064	0.795 ± 0.159
	CoModGan [69]			125.824	0.546 ± 0.230	10.652 ± 3.815	0.415 ± 0.306
	LaMa [50]			112.498	0.484 ± 0.229	11.152 ± 4.325	0.447 ± 0.272
Places	AIIA	Zeyu Lu	PyTorch	8.772	0.193 ± 0.209	24.145 ± 8.307	0.800 ± 0.188
	HSSLAB	Rengang Li	PyTorch	9.861	0.191 ± 0.217	24.345 ± 8.273	0.798 ± 0.191
	KwaiInpainting	Jiayin Cai	PyTorch	18.334	0.239 ± 0.193	23.410 ± 7.892	0.787 ± 0.195
	ArtificiallyInspired⭐	Ritwik Das	PyTorch & Tensorflow	7.544	0.204 ± 0.207	23.248 ± 9.477	0.777 ± 0.220
	SIGMA	Xiaoqiang Zhou	PyTorch	11.496	0.223 ± 0.189	22.562 ± 7.162	0.748 ± 0.200
	CoModGan [69]			67.910	0.496 ± 0.244	11.403 ± 4.154	0.477 ± 0.290
	LaMa [50]			66.566	0.523 ± 0.236	11.184 ± 3.758	0.328 ± 0.266
ImageNet	AIIA	Zeyu Lu	PyTorch	10.007	0.179 ± 0.216	25.226 ± 9.000	0.793 ± 0.214
	HSSLAB	Rengang Li	PyTorch	11.77	0.174 ± 0.219	24.303 ± 8.155	0.763 ± 0.221
	KwaiInpainting	Jiayin Cai	PyTorch	18.854	0.236 ± 0.243	23.804 ± 8.781	0.776 ± 0.221
	ArtificiallyInspired⭐	Ritwik Das	PyTorch & Tensorflow	12.059	0.196 ± 0.225	24.278 ± 9.814	0.779 ± 0.236
	SIGMA	Xiaoqiang Zhou	PyTorch	19.646	0.250 ± 0.203	22.454 ± 6.715	0.729 ± 0.227
	CoModGan [69]			40.826	0.450 ± 0.264	12.708 ± 5.366	0.526 ± 0.327
	LaMa [50]			42.204	0.504 ± 0.251	11.546 ± 4.080	0.260 ± 0.232
WikiArt	AIIA	Zeyu Lu	PyTorch	14.974	0.219 ± 0.215	24.350 ± 8.248	0.767 ± 0.203
	HSSLAB	Rengang Li	PyTorch	14.986	0.202 ± 0.187	24.257 ± 7.877	0.752 ± 0.206
	KwaiInpainting	Jiayin Cai	PyTorch	26.395	0.265 ± 0.212	23.142 ± 7.305	0.759 ± 0.204
	ArtificiallyInspired⭐	Ritwik Das	PyTorch & Tensorflow	8.524	0.229 ± 0.216	23.799 ± 9.230	0.758 ± 0.217
	SIGMA	Xiaoqiang Zhou	PyTorch	14.125	0.241 ± 0.192	22.717 ± 6.946	0.720 ± 0.208
	CoModGan [69]			89.117	0.513 ± 0.252	11.339 ± 4.514	0.477 ± 0.294
	LaMa [50]			88.473	0.552 ± 0.243	11.273 ± 4.191	0.297 ± 0.262

Table 2. Quantitative results over the Test set of Track 1.

	Team	MOS↑
FFHQ	AIIA	3.478
	HSSLAB	3.668
	ArtificiallyInspired⭐	4.503
	GT	4.953
Places	AIIA	3.675
	HSSLAB	3.561
	ArtificiallyInspired⭐	4.150
	GT	4.975
ImageNet	AIIA	3.986
	HSSLAB	4.046
	ArtificiallyInspired⭐	4.203
	GT	4.993
WikiArt	AIIA	3.653
	HSSLAB	3.607
	ArtificiallyInspired⭐	4.382
	GT	4.882

Table 3. Mean opinion score over the test set - Track 1.

Track 1 (CoModGAN [69]) by encoding the semantic information as a styled vector. In a more simple solution, *MGTV* concatenates the semantic information as input to the network.

Tables in Appendix C.2 depict the detailed results for each mask. Compared to Track 1, as the models leverage on the semantic information, the performance improves considerably. In general, the same behavior is presented in terms of the difficulty of the masks. However, the generation process is guided by exhibiting better perceptual results.

See additional qualitative results in Appendix B.2.

5. Challenge Methods

This section describes solutions submitted by all teams participating in the final stage of the NTIRE 2022 Image Inpainting Challenge.

5.1. Track 1

5.1.1 Artificially Inspired⭐

The proposed technique is based on CoModGAN [69]. As modulated convolutions are the core of the powerful StyleGAN [22], CoModGAN exploits them for the task of Image Inpainting. Please refer to the original paper for more details. The proposed solution is a fine-tuned

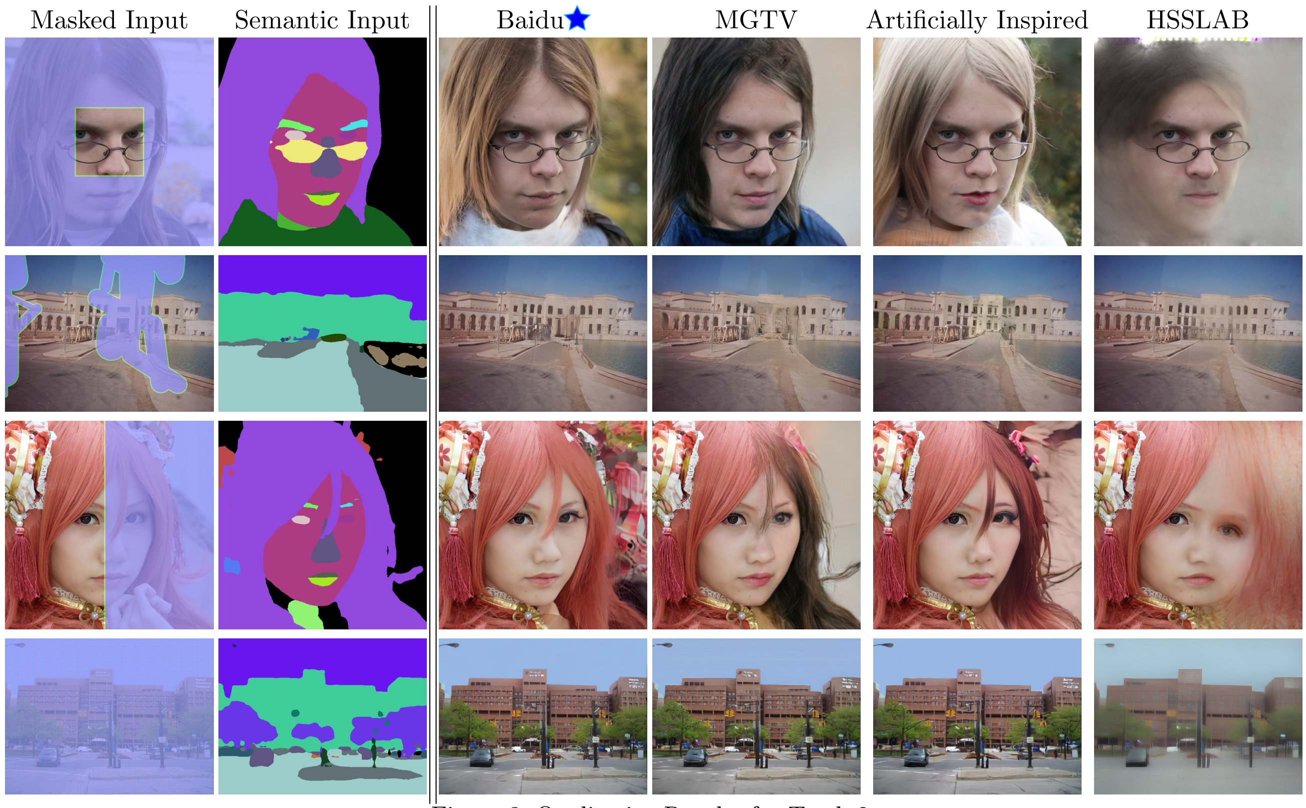

Figure 3. Qualitative Results for Track 2.

	Team	Author	Framework	FID↓	LPIPS↓	PSNR↑	SSIM↑	mIoU↑
FFHQ	MGTV	Xinying Wang	PyTorch	4.904	0.135 ± 0.132	25.701 ± 8.817	0.839 ± 0.143	0.962
	Baidu	Zhihong Pan	PyTorch	3.251	0.124 ± 0.133	26.156 ± 8.941	0.839 ± 0.143	0.963
	HSSLAB	Rengang Li	PyTorch	10.744	0.172 ± 0.186	26.882 ± 9.791	0.852 ± 0.146	0.842
	ArtificiallyInspired	Ritwik Das	PyTorch	3.662	0.139 ± 0.137	26.730 ± 9.855	0.840 ± 0.152	0.949
Places	MGTV	Xinying Wang	PyTorch	8.913	0.193 ± 0.180	23.740 ± 7.953	0.791 ± 0.191	0.668
	Baidu	Zhihong Pan	PyTorch	7.330	0.181 ± 0.188	23.330 ± 8.366	0.781 ± 0.201	0.635
	HSSLAB	Rengang Li	PyTorch	9.262	0.190 ± 0.216	25.739 ± 9.249	0.823 ± 0.189	0.592
	ArtificiallyInspired	Ritwik Das	PyTorch	7.295	0.188 ± 0.176	23.902 ± 8.940	0.784 ± 0.207	0.654

Table 4. Quantitative results over the Validation set of Track 2.

version of the publicly available CoModGAN models for the purpose of this challenge, namely with more generalized masks.

This method uses a sliding window technique to evaluate arbitrary-sized images during inference. If the input dimension is smaller than 512×512 (training dimension), both the image and the mask are padded accordingly.

Every dataset was trained on 8 V100 GPUs for two weeks. The runtime of this solution is roughly 1 second on average for the datasets evaluated.

Upon authorization from the Artificial Inspired team, we will release the inference code and pretrained weights in the challenge repo https://github.com/affromero/NTIRE22_Inpainting.

5.1.2 AIIA [27]

AIIA core solution adapts LaMA [51] for the current challenge. LaMa is a recent method that employs Fast Fourier Convolution [8] that allows a wide receptive field to perform inpainting efficiently and effectively. AIIA's proposed solution consists of two main improvements over the LaMa baseline. First, a focal frequency loss [20] that allows for a more narrow gap in the frequency domain. Second, in addition to the LaMa losses, it also employs a total variation loss [30] that helps to reduce the generated artifacts and chessboard. Figure 4 depicts the AIIA's proposed solution. Every

	Team	Author	Framework	FID↓	LPIPS↓	PSNR↑	SSIM↑	mIoU↑
FFHQ	MGTV	Xinying Wang	PyTorch	4.664	0.134 ± 0.131	25.769 ± 8.817	0.839 ± 0.143	0.962
	Baidu★	Zhihong Pan	PyTorch	3.193	0.123 ± 0.132	26.254 ± 8.892	0.838 ± 0.142	0.962
	HSSLAB	Rengang Li	PyTorch	13.504	0.171 ± 0.185	25.187 ± 8.864	0.821 ± 0.147	0.826
	ArtificiallyInspired	Ritwik Das	PyTorch	3.573	0.138 ± 0.136	26.827 ± 9.864	0.839 ± 0.153	0.948
Places	MGTV	Xinying Wang	PyTorch	8.735	0.193 ± 0.180	23.738 ± 7.838	0.791 ± 0.191	0.672
	Baidu★	Zhihong Pan	PyTorch	7.328	0.182 ± 0.188	23.289 ± 8.293	0.781 ± 0.201	0.636
	HSSLAB	Rengang Li	PyTorch	9.861	0.191 ± 0.217	24.345 ± 8.273	0.798 ± 0.191	0.574
	ArtificiallyInspired	Ritwik Das	PyTorch	7.229	0.188 ± 0.174	23.868 ± 8.864	0.784 ± 0.207	0.655

Table 5. Quantitative results over the Test set of Track 2.

	Team	MOS↑
FFHQ	MGTV	4.267
	Baidu★	4.582
	ArtificiallyInspired	4.575
	GT	4.900
Places	MGTV	3.546
	Baidu★	3.882
	ArtificiallyInspired	3.700
	GT	5

Table 6. Mean opinion score over the test set - Track 2.

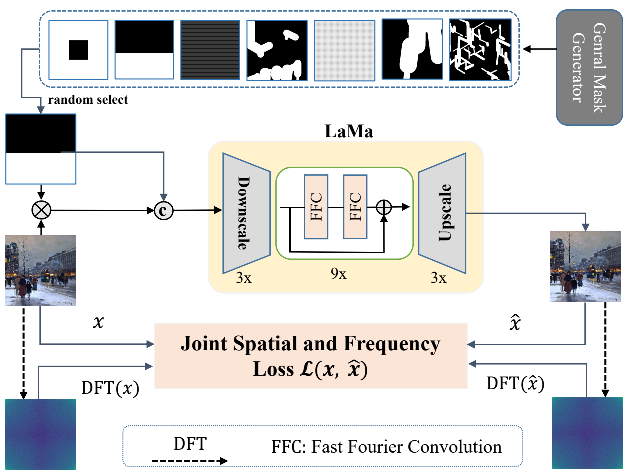

Figure 4. Diagram of AIIA. It employs an adaptation of LaMa [50].

dataset is trained on 8 V100 GPUs for three days.

5.1.3 HSSLAB

Similar to AIIA, the HSSLAB method relies on LaMa [51]. As the original LaMa only includes main strokes as a type of mask, the proposed technique includes the proposed challenge masks with additional data augmentation changes. Instead of randomly including the seven challenge masks, HSSLAB observation is to increase the probability of generation of *Nearest_Neighbor* masks, change the thickness of the stripes, add rectangular masks, and more out-painting samples with increasing difficulty during training. In total, this solution includes eight data augmentation techniques.

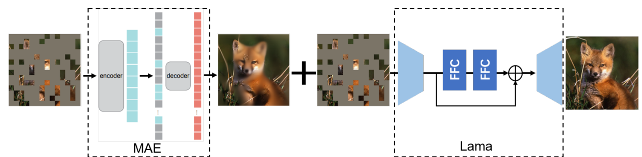

Figure 5. The solution proposed by SIGMA is an ensemble of a Masked AutoEncoder (MAE) [16] and LaMa [50]

5.1.4 SIGMA

SIGMA solution is an ensemble of a Masked Autoencoder (MAE) [16] and LaMa [51] (See Figure 5). It starts the process using MAE's transformer network to capture the long-range dependency in the image and learn the feature representation. Then, this complete image, along with the masked image, is passed on to LaMa to be used as a structure to guide the image synthesis and restore image details.

5.1.5 KwaiInpainting [52]

The core of this method relies on DeepFillv2 [60] with two main contributions: a progressive coarse-to-fine approach and a semantic-aware patchGAN. In addition to the GAN loss, it also employs the perceptual loss [66]. See Figure 6 for a schematic of the proposed solution.

5.2. Track 2

5.2.1 Baidu★

Team Baidu based their solution on diffusion models [48]. In detail, it relies on Latent Diffusion Models (LDM) [44], which conducts the diffusion and denoising process in a latent space of lower dimension. See the original paper for detailed information [44]. As shown in Figure 7, the Baidu solution for image inpainting

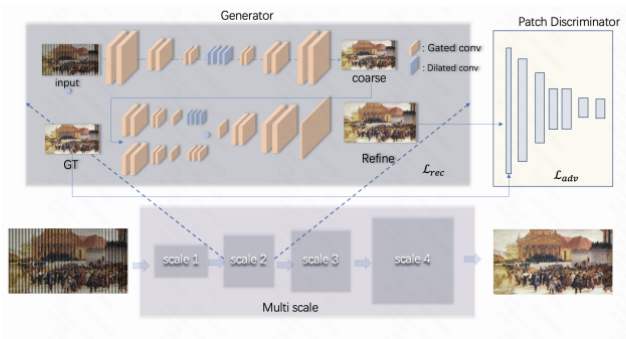

Figure 6. Diagram of the method proposed by KwaiInpainting team. It uses a multi-stage strategy and a semantic-aware discriminator.

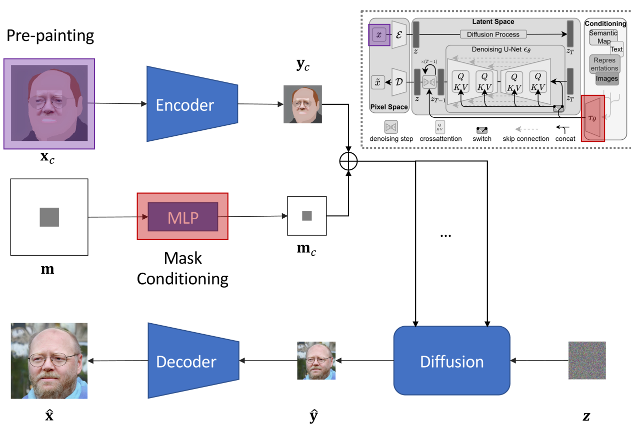

Figure 7. The solution proposed by Baidu consists of a Diffusion Model [44] guided by semantic information.

using semantic information consists of two key adaptations applied to LDM. First, pre-painting the masked image using semantic segmentation as a condition of the denoising process. Second, a mask conditioning MLP to break the limitation in resizing spatial masks like Nearest-Neighbor and Every-N-Lines.

There are a few techniques employed for performance improvement: 1) For Nearest-Neighbor and Every-N-Lines, the masked area of the input is pre-painted with bilinearly interpolated values without using semantic segmentation like other mask types; 2) At inference, 100 denoising steps are used for Completion and Expand, and 50 steps for all others; 3) Tone mapping is used to correct the color shift bias presented in predicted samples to match the color of the unmasked area of the input image; 4) For images larger than 512 $\times$ 512, a progressive sliding method is applied at inference for Completion and Expand masks to process the full image in 512 $\times$ 512 patches.

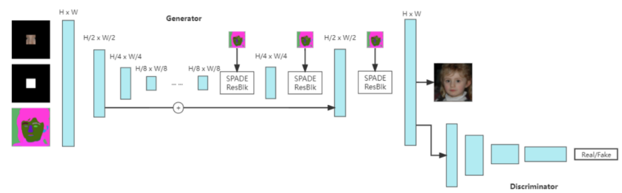

Figure 8. The solution presented by the MGTV team is a SPADE-based architecture [39]

5.2.2 Artificially Inspired

To include the semantic information, Artificially Inspired extended the solution proposed for Track 1 (Section 5.1.1). As CoModGAN uses modulated convolutions that depend on a *style vector*, in this track, this style vector is encoded from the semantic segmentation guidance. On the one hand, it has the advantage of using semantic information indirectly, and on the other hand, it converts CoModGAN into a fully-deterministic solution.

5.2.3 HSSLAB

HSSLAB proposes the same solution for both Track 1 and 2, and this method is described in Section 5.1.3.

5.2.4 MGTV

It uses a UNet-like architecture with SPADE normalization blocks to include the semantic information in the upsampling layers (Figure 8). SPADE normalizes the feature maps with respect to the semantic regions, and it has been well studied in semantic image synthesis [73, 5].

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).

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017. 1

[2] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral demosaicing challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[3] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral recovery challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[4] Goutam Bhat, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 burst super-resolution challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[5] Marcel Christoph Bühler, Andrés Romero, and Radu Timofte. DeepSEE: deep disentangled semantic explorative extreme super-resolution. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2020. 9

[6] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1209–1218, 2018. 3

[7] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 3

[8] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *Advances in Neural Information Processing Systems*, 33:4479–4488, 2020. 4, 7

[9] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021. 1, 4

[10] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. `https://github.com/open-mmlab/mmsegmentation`, 2020. 3

[11] 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, 3

[12] Egor Ershov, Alex Savchik, Denis Shepelev, Nikola Banic, Michael S Brown, Radu Timofte, et al. NTIRE 2022 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 1, 4

[14] Jinjin Gu, Haoming Cai, Chao Dong, Jimmy Ren, Radu Timofte, et al. NTIRE 2022 challenge on perceptual image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[15] Xiefan Guo, Hongyu Yang, and Di Huang. Image inpainting via conditional texture and structure dual generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14134–14143, 2021. 1

[16] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. *arXiv preprint arXiv:2111.06377*, 2021. 8

[17] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, 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. 3

[18] Seunghoon Hong, Xinchen Yan, Thomas Huang, and Honglak Lee. Learning hierarchical semantic image manipulation through structured representations. *arXiv preprint arXiv:1808.07535*, 2018. 1

[19] Zheng Hui, Jie Li, Xiumei Wang, and Xinbo Gao. Image fine-grained inpainting. *arXiv preprint arXiv:2002.02609*, 2020. 1

[20] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for generative models. *CoRR*, abs/2012.12821, 2020. 7

[21] 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. 1, 2, 3, 15

[22] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020. 4, 6

[23] Yawei Li, Kai Zhang, Radu Timofte, Luc Van Gool, et al. NTIRE 2022 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[24] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 1

[25] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018. 1

[26] Hongyu Liu, Bin Jiang, Yibing Song, Wei Huang, and Chao Yang. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 725–741. Springer, 2020. 1

[27] Zeyu Lu; Junjun Jiang ; Junqin Huang ; Gang Wu ; Xianming Liu. Glama: Joint spatial and frequency

loss for general image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 7

[28] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. *arXiv preprint arXiv:2201.09865*, 2022. 1, 4

[29] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 challenge on learning the super-resolution space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[30] Aravindh Mahendran and Andrea Vedaldi. Understanding deep image representations by inverting them. *CoRR*, abs/1412.0035, 2014. 7

[31] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021. 1, 4

[32] Pamela Mishkin, Lama Ahmad, Miles Brundage, Gretchen Krueger, and Girish Sastry. Dall·e 2 preview - risks and limitations. 2022. 4

[33] Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, and Kyoung Mu Lee. Ntire 2021 challenge on image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 149–165, 2021. 1

[34] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Qureshi, and Mehran Ebrahimi. Edgeconnect: Structure guided image inpainting using edge prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 1

[35] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021. 1, 4

[36] K. Nichol. Painter by numbers, wikiart. `https://www.kaggle.com/competitions/painter-by-numbers/`, 2016. 2, 3, 15

[37] Evangelos Ntavelis, Andrés Romero, Siavash Bigdeli, Radu Timofte, Zheng Hui, Xiumei Wang, Xinbo Gao, Chajin Shin, Taeoh Kim, Hanbin Son, et al. Aim 2020 challenge on image extreme inpainting. In *European Conference on Computer Vision*, pages 716–741. Springer, 2020. 1

[38] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis. In *European Conference on Computer Vision*, pages 739–755. Springer, 2020. 3

[39] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 9

[40] Jialun Peng, Dong Liu, Songcen Xu, and Houqiang Li. Generating diverse structure for image inpainting with hierarchical vq-vae. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10775–10784, 2021. 1

[41] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Richard Shaw, Ales Leonardis, Radu Timofte, et al. NTIRE 2022 challenge on high dynamic range imaging: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[42] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H Li, Shan Liu, and Ge Li. Structureflow: Image inpainting via structure-aware appearance flow. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 181–190, 2019. 1

[43] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2287–2296, 2021. 1

[44] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. *arXiv preprint arXiv:2112.10752*, 2021. 1, 4, 8, 9

[45] Andres Romero, Angela Castillo, Jose M Abril-Nova, Radu Timofte, et al. NTIRE 2022 image inpainting challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[46] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 15

[47] Chitwan Saharia, William Chan, Huiwen Chang, Chris A Lee, Jonathan Ho, Tim Salimans, David J Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. *arXiv preprint arXiv:2111.05826*, 2021. 1, 4

[48] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics, 2015. 8

[49] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations, 2020. 1, 4

[50] 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. *arXiv*

preprint arXiv:2109.07161, 2021. 1, 2, 4, 5, 6, 8, 16, 17, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29

[51] 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, 2021. 7, 8

[52] Jiayin Cai ; Changlin Li ; Xin Tao ; Yu-Wing Tai. Image multi-painting via progressive generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 8

[53] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 1

[54] Ziyu Wan, Jingbo Zhang, Dongdong Chen, and Jing Liao. High-fidelity pluralistic image completion with transformers. *arXiv preprint arXiv:2103.14031*, 2021. 1

[55] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Shuhang Gu, Radu Timofte, et al. NTIRE 2022 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[56] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018. 1

[57] Wei Xiong, Jiahui Yu, Zhe Lin, Jimei Yang, Xin Lu, Connelly Barnes, and Jiebo Luo. Foreground-aware image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5840–5848, 2019. 1

[58] Shunxin Xu, Dong Liu, and Zhiwei Xiong. E2i: Generative inpainting from edge to image. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(4):1308–1322, 2020. 1

[59] Ren Yang, Radu Timofte, et al. NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[60] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. *arXiv preprint arXiv:1801.07892*, 2018. 1, 4, 8

[61] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4471–4480, 2019. 1, 2

[62] Yingchen Yu, Fangneng Zhan, Rongliang Wu, Jianxiong Pan, Kaiwen Cui, Shijian Lu, Feiying Ma, Xuansong Xie, and Chunyan Miao. Diverse image inpainting with bidirectional and autoregressive transformers. In *Proceedings of the 29th ACM International Conference on Multimedia*, 2021. 1

[63] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Learning pyramid-context encoder network for high-quality image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1486–1494, 2019. 1

[64] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 1, 2

[65] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1671–1681, 2019. 1

[66] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *CVPR*, 2018. 3, 8

[67] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 1

[68] Lei Zhao, Qihang Mo, Sihuan Lin, Zhizhong Wang, Zhiwen Zuo, Haibo Chen, Wei Xing, and Dongming Lu. Uctgan: Diverse image inpainting based on unsupervised cross-space translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5741–5750, 2020. 1

[69] Shengyu Zhao, Jonathan Cui, Yilun Sheng, Yue Dong, Xiao Liang, Eric I Chang, and Yan Xu. Large scale image completion via co-modulated generative adversarial networks. *arXiv preprint arXiv:2103.10428*, 2021. 1, 2, 4, 5, 6, 16, 17, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29

[70] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. Pluralistic image completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1438–1447, 2019. 1

[71] 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. 2, 3

[72] 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*, 2017. 15

[73] Peihao Zhu, Rameen Abdal, Yipeng Qin, and Peter Wonka. Sean: Image synthesis with semantic region-adaptive normalization. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5104–5113, 2020. 9

NTIRE 2022 Image Inpainting Challenge: Report

Appendix

A. Teams and Affiliations

NTIRE 2022 Image Inpainting Team

Title:
NTIRE 2022 Image Inpainting Challenge
Members:
Andrés Romero[1] *(roandres@ethz.ch)*, Angela Castillo[2] *(a.castillo13@uniandes.edu.co)*, Jose Abril-Nova[2] *(jm.abril@uniandes.edu.co)*, Radu Timofte[3,1] *(radu.timofte@uni-wuerzburg.de)*
Affiliations:
[1] Computer Vision Lab, ETH Zürich, Switzerland
[2] Center for Research and Formation in Artificial Intelligence, Universidad de los Andes, Colombia
[3] University of Würzburg, Germany

Artificially Inspired

Title:
Image Inpainting using Comodulated Stylized Networks
Members:
Ritwik Das (ritwikdas54@gmail.com), Sanchit Hira
Affiliations:
Independent

Baidu

Title:
Semantic Image Inpainting using Learned Mask Condition in Latent Diffusion Model
Members:
Zhihong Pan (zhihongpan@baidu.com), Min Zhang, Baopu Li, Dongliang He, Tianwei Lin, Fu Li
Affiliations:
Baidu Research (USA)
University of Southern California
Department of Computer Vision Technology (VIS), Baidu Incorporation

AIIA LAB

Title:
GLaMa: A simple way to improve LaMa for general mask
Members:
Zeyu Lu (leo1037987031@gmail.com), Junqin Huang, Gang Wu, Junjun Jiang, Chengyue Wu, Xianming Liu
Affiliations:
Harbin Institute of Technology
Beihang University

MGTV

Title:
Image inpainting using semantic guidance with SPADE
Members:
Xinying Wang (xinying@mgtv.com), Yi Yu, Jie Yang
Affiliations:
MGTV

HSSLAB

Title:
LaMa for general masks
Members:
Rengang Li (lirengang.hsslab@gmail.com), Yaqian Zhao, Zhenhua Guo, Baoyu Fan, Xiaochuan Li, Runze Zhang
Affiliations:
Inspur Electronic Information Industry Co.,Ltd. and State Key Laboratory of High-end Server & Storage Technology

SIGMA

Title:
SIGMA: Solve Image Inpainting with Guidance from Masked Autoencoders
Members:
Xiaoqiang Zhou (13436433445@126.com)
Affiliations:
University of Science and Technology China

KwaiInpainting

Title:
multi-scale image inpainting network, in and out painting in one model, condi- tional projection patch discriminator.
Members:
Jiayin Cai (caijy18@tsinghua.org.cn), Changlin Li, XinTao, YuWing Tai
Affiliations:
Kuaishou

B. Additional Qualitative Results

B.1. Track 1: Unsupervised Image Inpainting

We show additional results for FFHQ, Places, ImageNet, and WikiArt in Figure 9, 10, 11, and 12, respectively.

B.2. Track 2: Image Inpainting guided by pixel-wise semantic labels

We show additional results for FFHQ and Places in Figure 13 and 14, respectively.

C. Additional Quantitative Results

C.1. Track 1: Unsupervised Image Inpainting

We show detailed **validation** scores per mask in Table 7, 8, 10, and 9, for FFHQ [21], Places [72], ImageNet [46], and WikiArt [36], respectively. Similarly, **test** scores in Table 11, 12, 14, and 13.

C.2. Track 2: Image Inpainting guided by pixel-wise semantic labels

We show detailed **validation** scores per mask in Table 15 and 16, for FFHQ and Places, respectively. Similarly, **test** scores in Table 17 and 18.

Figure 9. Additional Qualitative Results for Track 1 - FFHQ dataset.

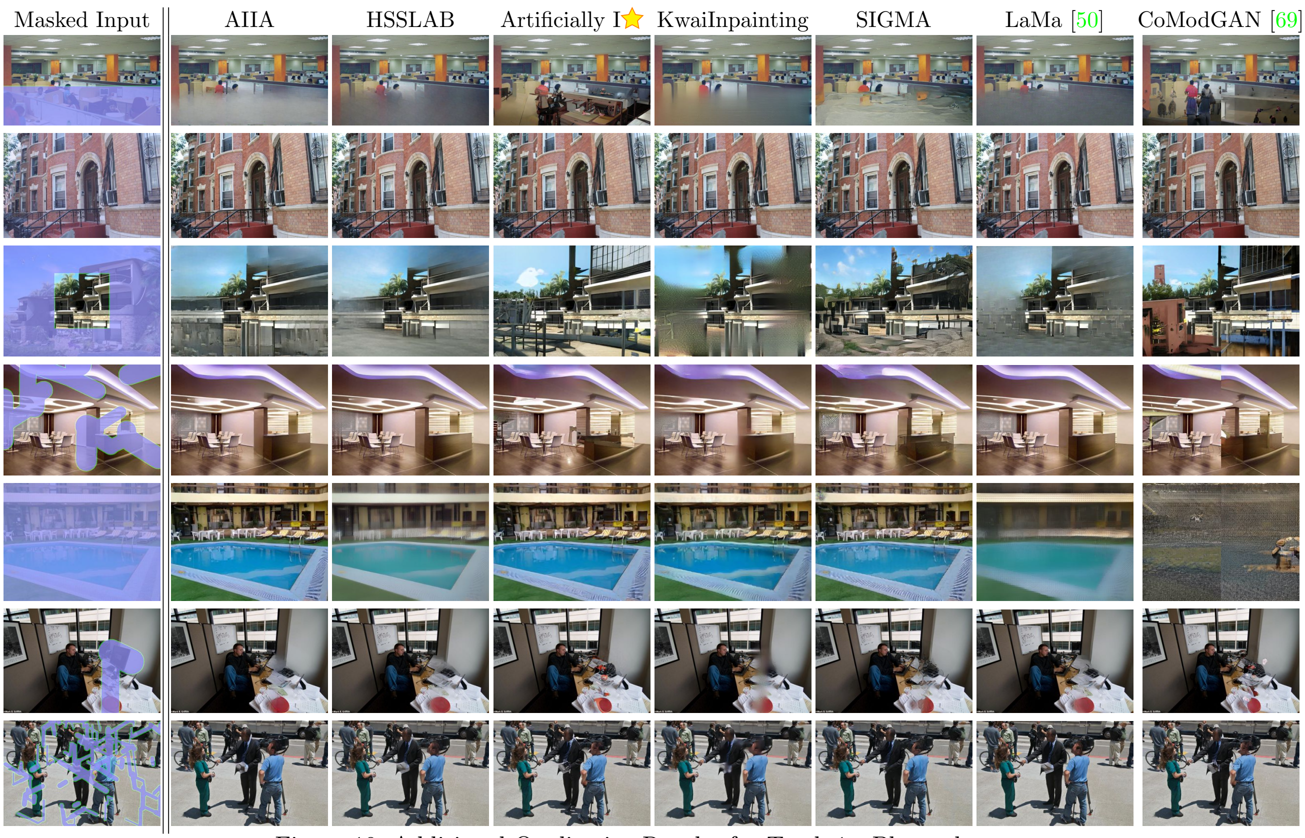

Figure 10. Additional Qualitative Results for Track 1 - Places dataset.

Figure 11. Additional Qualitative Results for Track 1 - ImageNet dataset.

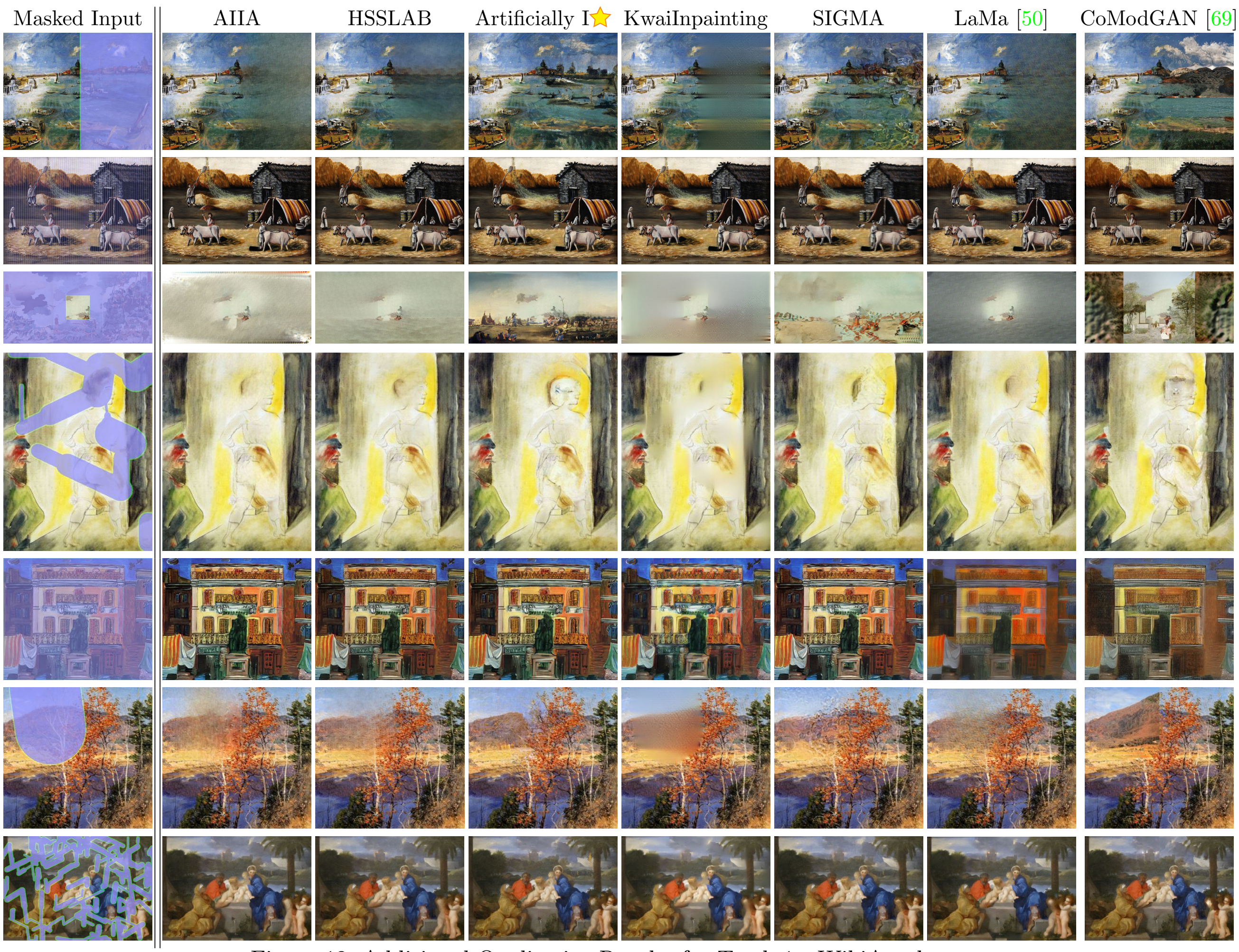

Figure 12. Additional Qualitative Results for Track 1 - WikiArt dataset.

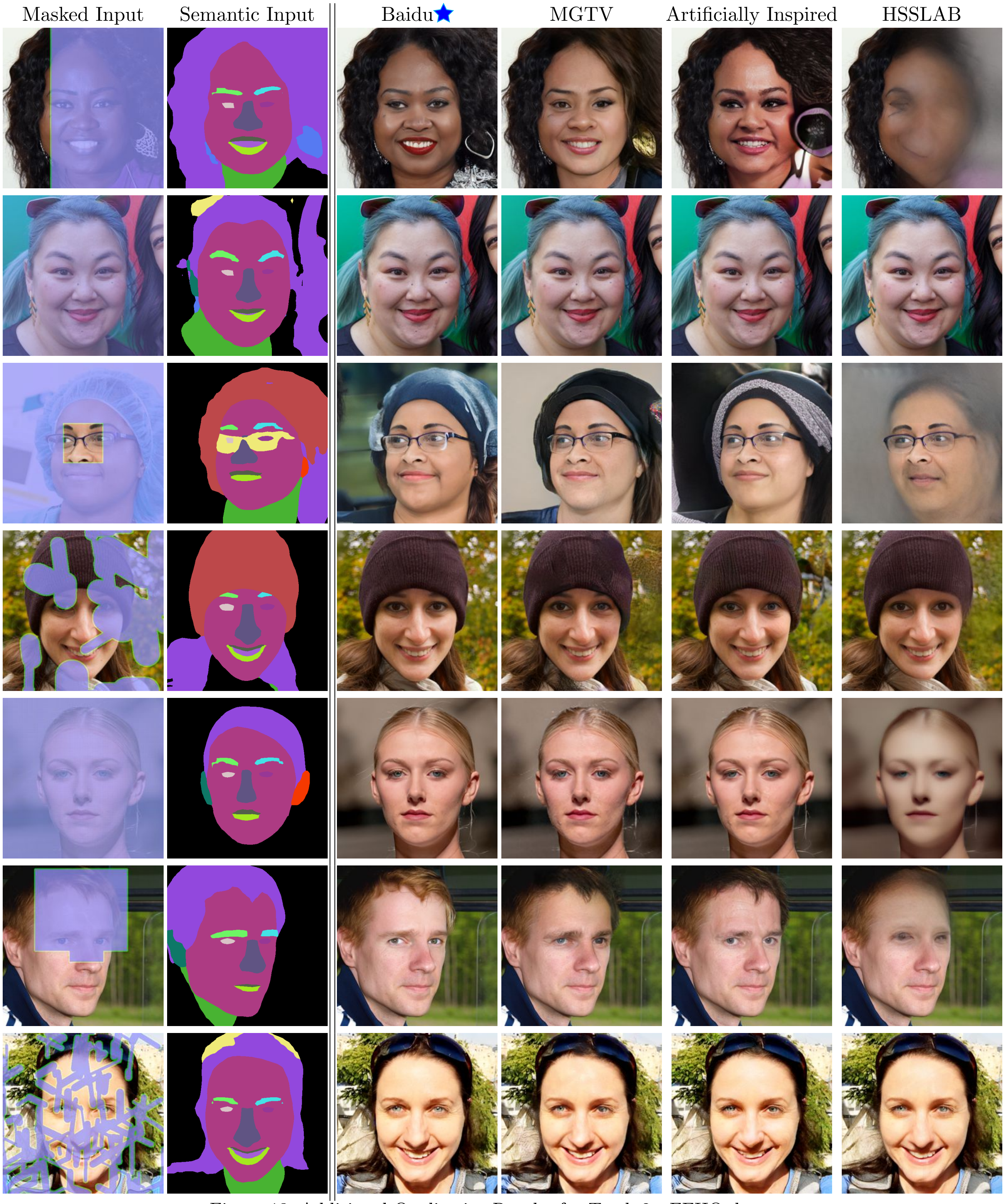

Figure 13. Additional Qualitative Results for Track 2 - FFHQ dataset.

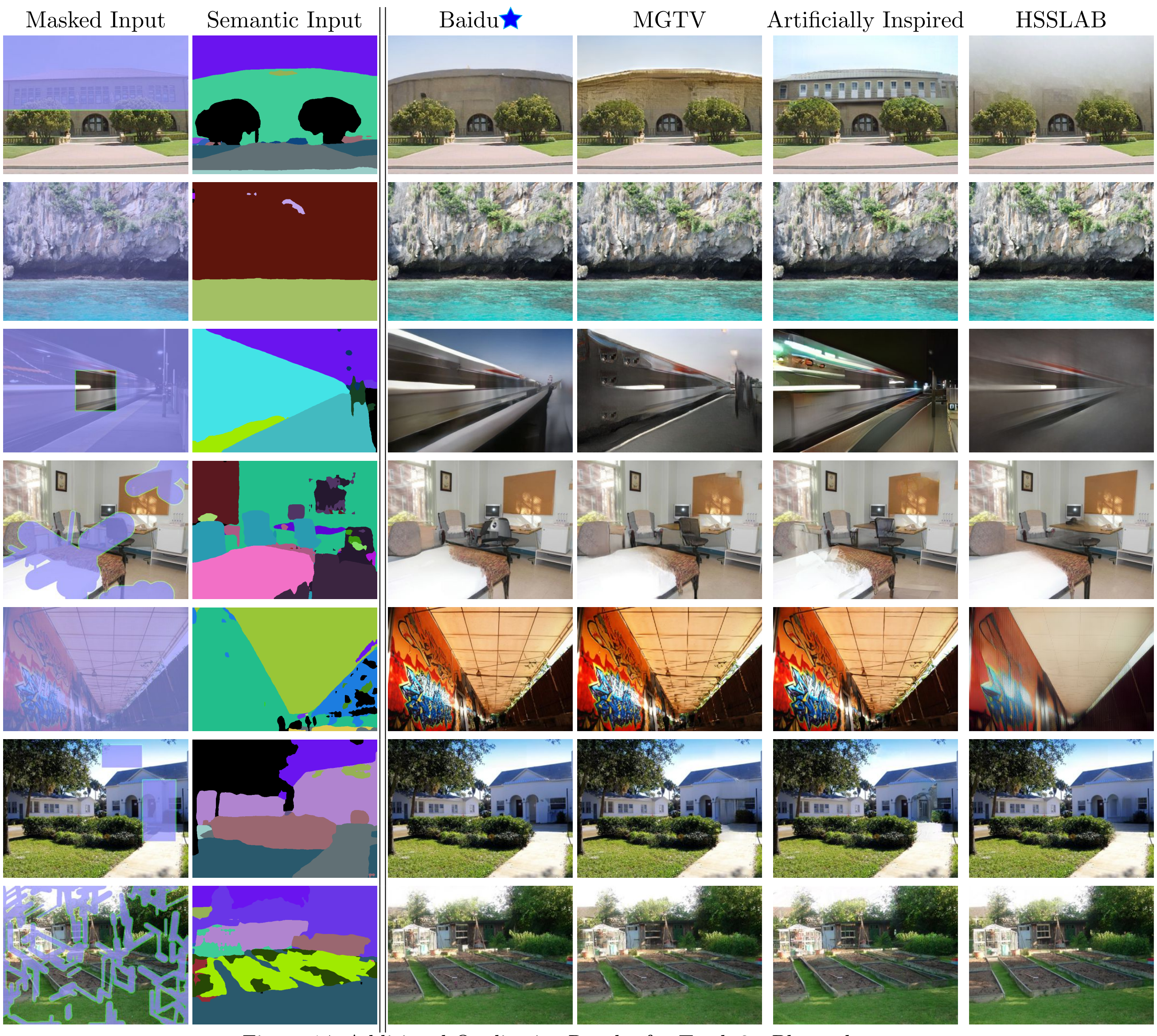

Figure 14. Additional Qualitative Results for Track 2 - Places dataset.

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑
FFHQ	Completion	AIIA	31.480	0.264 ± 0.113	17.270 ± 3.186	0.738 ± 0.107
		HSSLAB	32.675	0.271 ± 0.124	18.412 ± 3.534	0.786 ± 0.097
		KwaiInpainting	60.510	0.345 ± 0.139	16.732 ± 3.319	0.764 ± 0.103
		ArtificiallyInspired	21.156	0.255 ± 0.105	15.733 ± 3.434	0.718 ± 0.117
		SIGMA	25.306	0.243 ± 0.100	17.574 ± 3.446	0.739 ± 0.114
		CoModGan [69]	98.807	0.499 ± 0.183	9.509 ± 2.911	0.502 ± 0.172
		LaMa [50]	106.632	0.453 ± 0.179	10.220 ± 3.382	0.496 ± 0.172
	EveryNLines	AIIA	10.071	0.077 ± 0.054	35.678 ± 1.777	0.931 ± 0.022
		HSSLAB	5.615	0.031 ± 0.038	39.714 ± 4.572	0.959 ± 0.044
		KwaiInpainting	3.126	0.031 ± 0.018	35.788 ± 3.436	0.961 ± 0.018
		ArtificiallyInspired	1.083	0.011 ± 0.009	41.779 ± 2.975	0.984 ± 0.009
		SIGMA	8.773	0.068 ± 0.058	34.823 ± 2.263	0.909 ± 0.041
		CoModGan [69]	148.057	0.730 ± 0.091	10.901 ± 2.165	0.117 ± 0.153
		LaMa [50]	74.155	0.474 ± 0.231	13.926 ± 2.248	0.392 ± 0.143
	Expand	AIIA	102.086	0.526 ± 0.098	12.983 ± 2.145	0.511 ± 0.091
		HSSLAB	101.753	0.513 ± 0.092	13.681 ± 2.321	0.594 ± 0.092
		KwaiInpainting	180.345	0.613 ± 0.092	12.569 ± 2.197	0.554 ± 0.095
		ArtificiallyInspired	58.679	0.539 ± 0.103	10.876 ± 2.364	0.435 ± 0.100
		SIGMA	59.571	0.505 ± 0.078	12.402 ± 1.985	0.493 ± 0.090
		CoModGan [69]	279.175	0.779 ± 0.086	8.809 ± 2.550	0.152 ± 0.079
		LaMa [50]	336.515	0.728 ± 0.104	7.141 ± 2.253	0.130 ± 0.079
	MediumStrokes	AIIA	13.452	0.087 ± 0.050	26.409 ± 5.670	0.891 ± 0.061
		HSSLAB	12.837	0.075 ± 0.051	27.441 ± 6.805	0.921 ± 0.053
		KwaiInpainting	22.227	0.135 ± 0.073	25.716 ± 5.437	0.904 ± 0.055
		ArtificiallyInspired	11.857	0.083 ± 0.048	25.891 ± 6.025	0.886 ± 0.065
		SIGMA	15.195	0.094 ± 0.052	26.244 ± 5.604	0.886 ± 0.065
		CoModGan [69]	159.611	0.325 ± 0.141	12.598 ± 4.042	0.704 ± 0.145
		LaMa [50]	160.001	0.319 ± 0.130	12.886 ± 4.115	0.688 ± 0.150
	NearestNeighbor	AIIA	11.650	0.090 ± 0.040	29.575 ± 2.300	0.841 ± 0.053
		HSSLAB	24.288	0.176 ± 0.176	31.066 ± 6.309	0.843 ± 0.140
		KwaiInpainting	18.615	0.125 ± 0.058	30.273 ± 3.152	0.870 ± 0.051
		ArtificiallyInspired	14.360	0.110 ± 0.056	33.215 ± 3.485	0.908 ± 0.044
		SIGMA	17.571	0.170 ± 0.086	28.196 ± 2.493	0.754 ± 0.071
		CoModGan [69]	245.667	0.748 ± 0.092	7.111 ± 1.580	0.043 ± 0.034
		LaMa [50]	149.810	0.727 ± 0.169	7.597 ± 2.796	0.096 ± 0.082
	ThickStrokes	AIIA	13.953	0.106 ± 0.064	24.688 ± 6.116	0.881 ± 0.071
		HSSLAB	13.883	0.099 ± 0.066	25.191 ± 6.945	0.906 ± 0.062
		KwaiInpainting	24.624	0.157 ± 0.089	23.868 ± 6.121	0.894 ± 0.063
		ArtificiallyInspired	12.213	0.099 ± 0.063	24.071 ± 6.527	0.875 ± 0.075
		SIGMA	13.883	0.105 ± 0.063	24.962 ± 5.949	0.880 ± 0.072
		CoModGan [69]	118.506	0.291 ± 0.136	12.876 ± 4.271	0.727 ± 0.136
		LaMa [50]	115.686	0.271 ± 0.124	13.249 ± 4.455	0.718 ± 0.139
	ThinStrokes	AIIA	10.432	0.054 ± 0.030	30.385 ± 4.059	0.913 ± 0.048
		HSSLAB	9.042	0.039 ± 0.041	32.668 ± 5.754	0.954 ± 0.044
		KwaiInpainting	17.522	0.089 ± 0.048	30.019 ± 4.075	0.926 ± 0.041
		ArtificiallyInspired	9.740	0.055 ± 0.031	29.781 ± 4.249	0.909 ± 0.051
		SIGMA	14.767	0.070 ± 0.039	29.423 ± 4.111	0.903 ± 0.054
		CoModGan [69]	284.832	0.467 ± 0.158	12.069 ± 3.404	0.643 ± 0.159
		LaMa [50]	289.891	0.429 ± 0.127	12.349 ± 3.433	0.601 ± 0.172

Table 7. Detailed Validation Track 1 - FFHQ dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑
Places	Completion	AIIA	45.245	0.325 ± 0.157	16.651 ± 3.940	0.714 ± 0.139
		HSSLAB	45.657	0.313 ± 0.152	17.962 ± 4.278	0.735 ± 0.138
		KwaiInpainting	61.070	0.359 ± 0.153	16.480 ± 3.953	0.706 ± 0.143
		ArtificiallyInspired	45.429	0.331 ± 0.147	14.346 ± 3.960	0.654 ± 0.155
		SIGMA	47.469	0.299 ± 0.131	16.756 ± 4.089	0.681 ± 0.151
		CoModGan [69]	70.310	0.503 ± 0.188	10.355 ± 2.767	0.507 ± 0.170
		LaMa [50]	62.571	0.508 ± 0.176	11.424 ± 3.200	0.375 ± 0.170
	EveryNLines	AIIA	3.498	0.035 ± 0.027	34.464 ± 4.578	0.956 ± 0.032
		HSSLAB	3.999	0.027 ± 0.028	36.246 ± 4.628	0.964 ± 0.037
		KwaiInpainting	4.758	0.043 ± 0.032	32.471 ± 4.738	0.944 ± 0.038
		ArtificiallyInspired	2.413	0.022 ± 0.024	36.504 ± 5.134	0.973 ± 0.024
		SIGMA	15.010	0.141 ± 0.124	29.399 ± 4.352	0.867 ± 0.098
		CoModGan [69]	95.907	0.544 ± 0.136	12.873 ± 2.410	0.399 ± 0.158
		LaMa [50]	121.108	0.692 ± 0.129	10.901 ± 2.305	0.131 ± 0.118
	Expand	AIIA	93.091	0.604 ± 0.117	13.016 ± 2.595	0.456 ± 0.147
		HSSLAB	93.050	0.590 ± 0.124	13.968 ± 2.870	0.499 ± 0.150
		KwaiInpainting	135.283	0.655 ± 0.102	12.652 ± 2.626	0.439 ± 0.146
		ArtificiallyInspired	72.726	0.612 ± 0.098	10.688 ± 2.375	0.355 ± 0.136
		SIGMA	96.539	0.590 ± 0.098	12.439 ± 2.370	0.404 ± 0.136
		CoModGan [69]	244.386	0.880 ± 0.078	6.962 ± 1.796	0.114 ± 0.060
		LaMa [50]	210.674	0.885 ± 0.105	9.011 ± 2.721	0.070 ± 0.052
	MediumStrokes	AIIA	19.277	0.087 ± 0.057	25.624 ± 6.371	0.886 ± 0.079
		HSSLAB	17.306	0.075 ± 0.054	26.946 ± 7.173	0.910 ± 0.072
		KwaiInpainting	32.982	0.140 ± 0.084	25.315 ± 6.281	0.888 ± 0.078
		ArtificiallyInspired	21.832	0.105 ± 0.065	23.944 ± 6.423	0.869 ± 0.088
		SIGMA	23.096	0.106 ± 0.063	25.036 ± 5.990	0.872 ± 0.086
		CoModGan [69]	137.617	0.277 ± 0.141	13.638 ± 4.789	0.746 ± 0.136
		LaMa [50]	141.473	0.311 ± 0.137	13.067 ± 4.090	0.560 ± 0.196
	NearestNeighbor	AIIA	14.548	0.129 ± 0.069	26.508 ± 4.160	0.801 ± 0.116
		HSSLAB	25.402	0.182 ± 0.186	29.040 ± 7.111	0.819 ± 0.182
		KwaiInpainting	34.803	0.225 ± 0.112	24.896 ± 4.154	0.746 ± 0.134
		ArtificiallyInspired	18.713	0.165 ± 0.102	28.046 ± 4.987	0.841 ± 0.109
		SIGMA	23.556	0.227 ± 0.095	23.135 ± 3.555	0.656 ± 0.132
		CoModGan [69]	253.322	0.611 ± 0.133	8.667 ± 1.257	0.122 ± 0.091
		LaMa [50]	201.330	0.564 ± 0.093	7.531 ± 2.503	0.058 ± 0.078
	ThickStrokes	AIIA	23.017	0.117 ± 0.074	23.895 ± 6.903	0.869 ± 0.088
		HSSLAB	21.813	0.108 ± 0.073	24.926 ± 7.778	0.888 ± 0.082
		KwaiInpainting	35.425	0.165 ± 0.095	23.400 ± 6.925	0.869 ± 0.088
		ArtificiallyInspired	25.768	0.132 ± 0.078	21.979 ± 7.125	0.847 ± 0.098
		SIGMA	25.485	0.127 ± 0.073	23.526 ± 6.204	0.856 ± 0.095
		CoModGan [69]	103.779	0.272 ± 0.139	13.478 ± 4.749	0.740 ± 0.137
		LaMa [50]	107.666	0.302 ± 0.135	13.176 ± 4.234	0.561 ± 0.193
	ThinStrokes	AIIA	10.931	0.051 ± 0.031	28.836 ± 5.179	0.916 ± 0.058
		HSSLAB	8.536	0.036 ± 0.029	31.088 ± 6.101	0.947 ± 0.053
		KwaiInpainting	21.338	0.098 ± 0.060	28.243 ± 4.961	0.915 ± 0.057
		ArtificiallyInspired	14.722	0.071 ± 0.041	27.075 ± 5.235	0.899 ± 0.066
		SIGMA	18.041	0.080 ± 0.042	27.260 ± 4.832	0.895 ± 0.066
		CoModGan [69]	177.499	0.385 ± 0.157	13.277 ± 3.893	0.713 ± 0.132
		LaMa [50]	179.192	0.412 ± 0.153	12.717 ± 3.524	0.536 ± 0.192

Table 8. Detailed Validation Track 1 - Places dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑
WikiArt	Completion	AIIA	59.445	0.335 ± 0.150	17.821 ± 4.216	0.690 ± 0.146
		ArtificiallyInspired	46.046	0.333 ± 0.142	15.764 ± 4.160	0.648 ± 0.157
		HSSLAB	57.208	0.308 ± 0.142	19.032 ± 4.304	0.717 ± 0.145
		KwaiInpainting	80.706	0.365 ± 0.148	17.555 ± 4.114	0.696 ± 0.149
		SIGMA	51.591	0.306 ± 0.130	17.616 ± 4.147	0.667 ± 0.154
		CoModGan [69]	109.840	0.519 ± 0.187	10.097 ± 2.724	0.500 ± 0.170
		LaMa [50]	96.648	0.537 ± 0.181	11.272 ± 3.250	0.339 ± 0.180
	EveryNLines	AIIA	7.142	0.049 ± 0.044	33.988 ± 4.882	0.930 ± 0.054
		ArtificiallyInspired	5.378	0.035 ± 0.035	36.031 ± 5.560	0.954 ± 0.039
		HSSLAB	9.266	0.049 ± 0.050	35.880 ± 4.859	0.943 ± 0.046
		KwaiInpainting	9.663	0.063 ± 0.047	30.550 ± 4.686	0.911 ± 0.056
		SIGMA	25.096	0.166 ± 0.105	28.085 ± 4.269	0.824 ± 0.092
		CoModGan [69]	120.675	0.571 ± 0.161	12.950 ± 3.065	0.399 ± 0.181
		LaMa [50]	149.480	0.728 ± 0.141	10.924 ± 3.066	0.097 ± 0.122
	Expand	AIIA	140.356	0.646 ± 0.116	12.877 ± 3.285	0.413 ± 0.166
		ArtificiallyInspired	70.428	0.639 ± 0.099	11.617 ± 2.903	0.364 ± 0.140
		HSSLAB	112.234	0.589 ± 0.125	14.680 ± 3.535	0.475 ± 0.171
		KwaiInpainting	150.421	0.660 ± 0.105	13.287 ± 3.188	0.426 ± 0.168
		SIGMA	106.047	0.611 ± 0.102	13.327 ± 3.181	0.398 ± 0.161
		CoModGan [69]	308.476	0.901 ± 0.076	6.416 ± 1.655	0.107 ± 0.062
		LaMa [50]	275.811	0.908 ± 0.099	9.024 ± 2.854	0.065 ± 0.052
	MediumStrokes	AIIA	28.644	0.108 ± 0.066	25.844 ± 5.894	0.868 ± 0.087
		ArtificiallyInspired	26.107	0.116 ± 0.069	24.507 ± 6.086	0.855 ± 0.094
		HSSLAB	24.665	0.089 ± 0.061	26.960 ± 6.563	0.899 ± 0.079
		KwaiInpainting	50.474	0.162 ± 0.091	25.041 ± 5.532	0.876 ± 0.082
		SIGMA	29.956	0.118 ± 0.068	25.269 ± 5.772	0.859 ± 0.092
		CoModGan [69]	172.011	0.293 ± 0.146	13.606 ± 4.885	0.746 ± 0.135
		LaMa [50]	175.952	0.339 ± 0.144	12.980 ± 4.411	0.510 ± 0.215
	NearestNeighbor	AIIA	24.876	0.189 ± 0.106	26.381 ± 4.592	0.724 ± 0.146
		ArtificiallyInspired	32.178	0.250 ± 0.144	27.755 ± 5.417	0.774 ± 0.139
		HSSLAB	39.325	0.207 ± 0.191	27.314 ± 6.713	0.750 ± 0.215
		KwaiInpainting	49.009	0.298 ± 0.133	23.511 ± 4.082	0.644 ± 0.156
		SIGMA	34.909	0.256 ± 0.104	22.602 ± 3.845	0.569 ± 0.146
		CoModGan [69]	232.360	0.609 ± 0.171	8.491 ± 1.618	0.135 ± 0.137
		LaMa [50]	191.660	0.602 ± 0.117	7.464 ± 3.258	0.045 ± 0.079
	ThickStrokes	AIIA	33.025	0.142 ± 0.088	23.949 ± 6.366	0.848 ± 0.097
		ArtificiallyInspired	28.948	0.144 ± 0.085	22.494 ± 6.591	0.831 ± 0.106
		HSSLAB	29.278	0.122 ± 0.081	24.846 ± 7.039	0.872 ± 0.091
		KwaiInpainting	50.202	0.186 ± 0.104	23.009 ± 6.197	0.852 ± 0.096
		SIGMA	31.063	0.141 ± 0.082	23.636 ± 6.130	0.838 ± 0.103
		CoModGan [69]	128.603	0.286 ± 0.144	13.248 ± 4.696	0.736 ± 0.138
		LaMa [50]	135.064	0.328 ± 0.140	12.945 ± 4.448	0.506 ± 0.207
	ThinStrokes	AIIA	19.859	0.073 ± 0.043	28.566 ± 5.271	0.888 ± 0.072
		ArtificiallyInspired	21.117	0.088 ± 0.050	27.174 ± 5.513	0.873 ± 0.081
		HSSLAB	16.144	0.053 ± 0.042	30.773 ± 6.363	0.931 ± 0.066
		KwaiInpainting	38.984	0.130 ± 0.077	27.525 ± 4.851	0.892 ± 0.071
		SIGMA	27.601	0.100 ± 0.056	27.003 ± 5.055	0.870 ± 0.081
		CoModGan [69]	218.209	0.418 ± 0.174	13.147 ± 4.252	0.706 ± 0.143
		LaMa [50]	218.445	0.452 ± 0.170	12.543 ± 3.906	0.479 ± 0.206

Table 9. Detailed Validation Track 1 - Wikiart dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑
ImageNet	Completion	AIIA	52.467	0.326 ± 0.156	16.848 ± 4.298	0.685 ± 0.149
		HSSLAB	49.421	0.300 ± 0.149	18.180 ± 4.546	0.709 ± 0.147
		KwaiInpainting	65.015	0.372 ± 0.160	16.111 ± 4.143	0.683 ± 0.150
		ArtificiallyInspired	64.521	0.335 ± 0.149	14.539 ± 4.009	0.637 ± 0.157
		SIGMA	64.697	0.332 ± 0.135	16.905 ± 3.871	0.661 ± 0.152
		CoModGan [69]	67.921	0.493 ± 0.186	10.213 ± 3.216	0.509 ± 0.168
		LaMa [50]	60.158	0.538 ± 0.183	11.092 ± 3.280	0.269 ± 0.151
	EveryNLines	AIIA	4.057	0.038 ± 0.029	32.060 ± 5.098	0.932 ± 0.050
		HSSLAB	4.005	0.034 ± 0.037	34.065 ± 5.446	0.945 ± 0.052
		KwaiInpainting	6.989	0.058 ± 0.041	29.854 ± 4.762	0.914 ± 0.055
		ArtificiallyInspired	3.218	0.030 ± 0.029	33.657 ± 5.409	0.951 ± 0.040
		SIGMA	26.220	0.200 ± 0.114	25.851 ± 3.371	0.804 ± 0.099
		CoModGan [69]	85.583	0.539 ± 0.152	13.145 ± 2.751	0.425 ± 0.172
		LaMa [50]	111.663	0.706 ± 0.137	10.759 ± 2.585	0.114 ± 0.118
	Expand	AIIA	110.120	0.601 ± 0.126	12.980 ± 2.942	0.409 ± 0.173
		HSSLAB	92.768	0.566 ± 0.135	13.733 ± 3.254	0.450 ± 0.179
		KwaiInpainting	137.069	0.684 ± 0.129	11.856 ± 2.857	0.402 ± 0.165
		ArtificiallyInspired	122.694	0.635 ± 0.107	10.391 ± 2.422	0.328 ± 0.144
		SIGMA	143.855	0.628 ± 0.107	12.367 ± 2.521	0.374 ± 0.153
		CoModGan [69]	192.334	0.846 ± 0.096	7.540 ± 2.193	0.117 ± 0.068
		LaMa [50]	172.496	0.869 ± 0.116	8.974 ± 3.145	0.051 ± 0.048
	MediumStrokes	AIIA	6.410	0.034 ± 0.028	31.970 ± 8.582	0.948 ± 0.044
		HSSLAB	5.014	0.026 ± 0.024	34.188 ± 10.276	0.963 ± 0.038
		KwaiInpainting	10.672	0.062 ± 0.049	30.945 ± 8.998	0.947 ± 0.046
		ArtificiallyInspired	7.654	0.041 ± 0.032	30.588 ± 8.820	0.941 ± 0.048
		SIGMA	13.687	0.081 ± 0.036	28.210 ± 4.052	0.930 ± 0.049
		CoModGan [69]	60.368	0.162 ± 0.096	18.388 ± 6.371	0.883 ± 0.076
		LaMa [50]	64.712	0.237 ± 0.095	14.896 ± 3.616	0.461 ± 0.206
	NearestNeighbor	AIIA	19.938	0.152 ± 0.094	25.011 ± 4.772	0.728 ± 0.153
		HSSLAB	31.189	0.193 ± 0.192	27.799 ± 6.729	0.774 ± 0.209
		KwaiInpainting	62.126	0.309 ± 0.154	22.421 ± 4.168	0.645 ± 0.165
		ArtificiallyInspired	29.493	0.206 ± 0.133	26.088 ± 5.500	0.772 ± 0.152
		SIGMA	42.165	0.325 ± 0.126	20.595 ± 3.319	0.525 ± 0.143
		CoModGan [69]	178.784	0.617 ± 0.148	8.683 ± 1.542	0.121 ± 0.109
		LaMa [50]	151.271	0.571 ± 0.108	7.860 ± 2.738	0.054 ± 0.079
	ThickStrokes	AIIA	17.338	0.086 ± 0.066	26.012 ± 6.974	0.886 ± 0.088
		HSSLAB	13.799	0.071 ± 0.059	27.379 ± 8.012	0.910 ± 0.078
		KwaiInpainting	25.386	0.131 ± 0.091	25.101 ± 6.818	0.886 ± 0.087
		ArtificiallyInspired	20.256	0.097 ± 0.071	24.580 ± 7.118	0.873 ± 0.096
		SIGMA	24.613	0.132 ± 0.069	24.604 ± 4.748	0.866 ± 0.094
		CoModGan [69]	74.570	0.237 ± 0.143	14.929 ± 5.356	0.787 ± 0.135
		LaMa [50]	80.411	0.303 ± 0.137	13.224 ± 3.765	0.420 ± 0.196
	ThinStrokes	AIIA	4.928	0.027 ± 0.020	32.128 ± 5.773	0.950 ± 0.040
		HSSLAB	3.409	0.017 ± 0.017	35.328 ± 7.259	0.972 ± 0.033
		KwaiInpainting	8.532	0.045 ± 0.035	31.099 ± 5.307	0.949 ± 0.041
		ArtificiallyInspired	7.095	0.037 ± 0.027	30.563 ± 5.915	0.941 ± 0.046
		SIGMA	15.244	0.080 ± 0.036	28.276 ± 3.588	0.922 ± 0.052
		CoModGan [69]	112.792	0.287 ± 0.135	16.453 ± 4.170	0.833 ± 0.092
		LaMa [50]	114.879	0.341 ± 0.129	14.296 ± 3.272	0.447 ± 0.195

Table 10. Detailed Validation Track 1 - ImageNet dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑
FFHQ	Completion	AIIA	31.345	0.295 ± 0.112	17.212 ± 3.366	0.735 ± 0.108
		HSSLAB	33.486	0.300 ± 0.121	17.128 ± 3.665	0.757 ± 0.101
		KwaiInpainting	60.420	0.322 ± 0.125	16.702 ± 3.485	0.763 ± 0.102
		ArtificiallyInspired	21.663	0.291 ± 0.110	15.857 ± 3.665	0.716 ± 0.117
		SIGMA	25.226	0.284 ± 0.105	17.592 ± 3.520	0.736 ± 0.113
		CoModGan [69]	98.368	0.498 ± 0.184	9.713 ± 3.129	0.502 ± 0.172
		LaMa [50]	106.641	0.451 ± 0.180	10.465 ± 3.669	0.497 ± 0.173
	EveryNLines	AIIA	9.853	0.190 ± 0.051	35.747 ± 1.894	0.931 ± 0.023
		HSSLAB	6.055	0.087 ± 0.055	37.019 ± 4.068	0.958 ± 0.027
		KwaiInpainting	3.009	0.103 ± 0.048	35.969 ± 3.512	0.961 ± 0.018
		ArtificiallyInspired	1.011	0.032 ± 0.019	41.778 ± 2.975	0.984 ± 0.009
		SIGMA	8.437	0.192 ± 0.075	34.874 ± 2.369	0.910 ± 0.043
		CoModGan [69]	146.304	0.719 ± 0.093	11.063 ± 2.256	0.127 ± 0.156
		LaMa [50]	75.865	0.469 ± 0.229	14.102 ± 2.360	0.400 ± 0.144
	Expand	AIIA	101.904	0.554 ± 0.082	13.028 ± 2.264	0.509 ± 0.091
		HSSLAB	103.304	0.536 ± 0.077	12.750 ± 2.227	0.563 ± 0.093
		KwaiInpainting	180.148	0.574 ± 0.077	12.480 ± 2.204	0.551 ± 0.097
		ArtificiallyInspired	59.827	0.567 ± 0.083	10.985 ± 2.484	0.437 ± 0.103
		SIGMA	59.769	0.546 ± 0.071	12.447 ± 2.046	0.494 ± 0.093
		CoModGan [69]	278.421	0.778 ± 0.086	8.813 ± 2.607	0.153 ± 0.081
		LaMa [50]	333.231	0.728 ± 0.104	7.120 ± 2.346	0.131 ± 0.081
	MediumStrokes	AIIA	12.830	0.129 ± 0.068	26.525 ± 5.736	0.891 ± 0.063
		HSSLAB	17.545	0.137 ± 0.087	26.147 ± 5.730	0.884 ± 0.072
		KwaiInpainting	21.740	0.152 ± 0.078	25.894 ± 5.562	0.904 ± 0.056
		ArtificiallyInspired	11.386	0.123 ± 0.065	26.038 ± 6.103	0.885 ± 0.067
		SIGMA	14.571	0.138 ± 0.070	26.371 ± 5.567	0.886 ± 0.066
		CoModGan [69]	164.666	0.324 ± 0.141	12.672 ± 4.410	0.704 ± 0.145
		LaMa [50]	164.672	0.319 ± 0.130	12.938 ± 4.471	0.688 ± 0.151
	NearestNeighbor	AIIA	12.067	0.248 ± 0.046	29.490 ± 2.376	0.837 ± 0.057
		HSSLAB	39.959	0.335 ± 0.172	29.002 ± 4.705	0.798 ± 0.110
		KwaiInpainting	19.571	0.230 ± 0.077	30.262 ± 3.150	0.869 ± 0.054
		ArtificiallyInspired	14.966	0.178 ± 0.067	33.159 ± 3.556	0.907 ± 0.046
		SIGMA	18.374	0.301 ± 0.058	28.193 ± 2.608	0.753 ± 0.074
		CoModGan [69]	242.698	0.746 ± 0.091	7.205 ± 1.585	0.046 ± 0.036
		LaMa [50]	149.602	0.725 ± 0.170	7.694 ± 2.779	0.094 ± 0.079
	ThickStrokes	AIIA	13.428	0.134 ± 0.074	24.765 ± 5.997	0.882 ± 0.071
		HSSLAB	17.633	0.145 ± 0.086	24.150 ± 5.807	0.878 ± 0.072
		KwaiInpainting	25.109	0.156 ± 0.082	23.992 ± 5.967	0.895 ± 0.064
		ArtificiallyInspired	11.605	0.128 ± 0.073	24.269 ± 6.433	0.877 ± 0.075
		SIGMA	13.601	0.137 ± 0.074	25.010 ± 5.898	0.881 ± 0.073
		CoModGan [69]	119.135	0.290 ± 0.137	12.892 ± 4.317	0.727 ± 0.136
		LaMa [50]	116.625	0.271 ± 0.124	13.258 ± 4.475	0.718 ± 0.139
	ThinStrokes	AIIA	9.983	0.119 ± 0.057	30.449 ± 4.175	0.912 ± 0.049
		HSSLAB	12.709	0.111 ± 0.074	30.113 ± 5.090	0.908 ± 0.059
		KwaiInpainting	17.384	0.136 ± 0.066	30.123 ± 4.320	0.925 ± 0.042
		ArtificiallyInspired	9.190	0.116 ± 0.055	29.906 ± 4.418	0.908 ± 0.052
		SIGMA	14.007	0.142 ± 0.067	29.535 ± 4.239	0.902 ± 0.055
		CoModGan [69]	282.020	0.463 ± 0.158	12.209 ± 3.431	0.645 ± 0.158
		LaMa [50]	285.306	0.426 ± 0.128	12.486 ± 3.448	0.603 ± 0.171

Table 11. Detailed Test Track 1 - FFHQ dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑
Places	Completion	AIIA	47.357	0.326 ± 0.136	16.588 ± 4.030	0.701 ± 0.142
		HSSLAB	46.38	0.316 ± 0.130	16.921 ± 4.244	0.705 ± 0.144
		KwaiInpainting	62.936	0.337 ± 0.132	16.513 ± 3.958	0.693 ± 0.148
		ArtificiallyInspired	46.091	0.338 ± 0.137	14.277 ± 3.967	0.641 ± 0.159
		SIGMA	49.437	0.321 ± 0.125	16.708 ± 4.006	0.668 ± 0.155
		CoModGan [69]	72.826	0.514 ± 0.193	10.491 ± 2.814	0.498 ± 0.175
		LaMa [50]	63.599	0.516 ± 0.181	11.565 ± 3.175	0.368 ± 0.175
	EveryNLines	AIIA	3.409	0.091 ± 0.050	34.463 ± 4.554	0.956 ± 0.036
		HSSLAB	4.772	0.086 ± 0.067	34.130 ± 4.544	0.956 ± 0.038
		KwaiInpainting	4.548	0.100 ± 0.049	32.516 ± 4.485	0.946 ± 0.040
		ArtificiallyInspired	2.301	0.041 ± 0.031	36.474 ± 5.112	0.973 ± 0.025
		SIGMA	13.768	0.220 ± 0.102	29.541 ± 4.223	0.877 ± 0.084
		CoModGan [69]	96.899	0.550 ± 0.141	12.944 ± 2.435	0.398 ± 0.160
		LaMa [50]	120.704	0.690 ± 0.126	10.976 ± 2.374	0.135 ± 0.118
	Expand	AIIA	90.619	0.593 ± 0.089	13.104 ± 2.574	0.463 ± 0.149
		HSSLAB	87.303	0.572 ± 0.092	13.244 ± 2.782	0.483 ± 0.152
		KwaiInpainting	135.921	0.603 ± 0.084	12.768 ± 2.587	0.445 ± 0.150
		ArtificiallyInspired	72.053	0.607 ± 0.082	10.878 ± 2.241	0.363 ± 0.135
		SIGMA	96.439	0.594 ± 0.078	12.543 ± 2.388	0.412 ± 0.139
		CoModGan [69]	242.577	0.875 ± 0.079	6.968 ± 1.785	0.116 ± 0.059
		LaMa [50]	205.250	0.879 ± 0.100	9.055 ± 2.676	0.071 ± 0.048
	MediumStrokes	AIIA	17.812	0.110 ± 0.064	25.660 ± 6.236	0.888 ± 0.078
		HSSLAB	16.746	0.108 ± 0.059	25.878 ± 5.307	0.888 ± 0.075
		KwaiInpainting	30.531	0.138 ± 0.077	25.330 ± 6.021	0.890 ± 0.075
		ArtificiallyInspired	20.307	0.126 ± 0.072	24.008 ± 6.256	0.871 ± 0.086
		SIGMA	21.742	0.132 ± 0.071	25.109 ± 5.862	0.874 ± 0.084
		CoModGan [69]	134.389	0.273 ± 0.138	13.709 ± 4.767	0.750 ± 0.134
		LaMa [50]	137.856	0.307 ± 0.132	13.081 ± 3.850	0.561 ± 0.189
	NearestNeighbor	AIIA	14.160	0.228 ± 0.065	26.674 ± 4.047	0.805 ± 0.113
		HSSLAB	37.341	0.291 ± 0.183	26.998 ± 6.276	0.765 ± 0.184
		KwaiInpainting	34.299	0.336 ± 0.122	24.987 ± 4.013	0.749 ± 0.132
		ArtificiallyInspired	18.421	0.203 ± 0.092	28.213 ± 4.825	0.845 ± 0.107
		SIGMA	23.253	0.348 ± 0.082	23.211 ± 3.414	0.655 ± 0.126
		CoModGan [69]	249.782	0.609 ± 0.138	8.761 ± 1.291	0.121 ± 0.094
		LaMa [50]	200.398	0.563 ± 0.091	7.579 ± 2.504	0.052 ± 0.075
	ThickStrokes	AIIA	23.031	0.134 ± 0.077	23.440 ± 6.319	0.866 ± 0.088
		HSSLAB	22.035	0.132 ± 0.073	23.740 ± 5.687	0.867 ± 0.086
		KwaiInpainting	35.220	0.156 ± 0.086	23.187 ± 6.397	0.867 ± 0.087
		ArtificiallyInspired	24.899	0.148 ± 0.083	21.563 ± 6.350	0.844 ± 0.098
		SIGMA	25.246	0.148 ± 0.080	23.252 ± 5.903	0.853 ± 0.094
		CoModGan [69]	101.557	0.271 ± 0.137	13.549 ± 4.600	0.739 ± 0.137
		LaMa [50]	105.310	0.300 ± 0.132	13.186 ± 3.919	0.558 ± 0.188
	ThinStrokes	AIIA	10.605	0.087 ± 0.046	29.085 ± 5.013	0.920 ± 0.056
		HSSLAB	9.804	0.081 ± 0.040	29.503 ± 4.723	0.920 ± 0.054
		KwaiInpainting	20.601	0.116 ± 0.061	28.572 ± 4.856	0.919 ± 0.056
		ArtificiallyInspired	14.029	0.111 ± 0.056	27.320 ± 5.100	0.903 ± 0.065
		SIGMA	17.196	0.126 ± 0.061	27.570 ± 4.772	0.899 ± 0.065
		CoModGan [69]	177.811	0.380 ± 0.154	13.398 ± 3.803	0.720 ± 0.134
		LaMa [50]	178.810	0.406 ± 0.149	12.845 ± 3.386	0.550 ± 0.187

Table 12. Detailed Test Track 1 - Places dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑
Wikiart	Completion	AIIA	59.04	0.314 ± 0.121	17.871 ± 4.427	0.696 ± 0.143
		HSSLAB	54.893	0.309 ± 0.117	18.143 ± 4.555	0.695 ± 0.143
		KwaiInpainting	79.088	0.326 ± 0.123	17.658 ± 4.411	0.702 ± 0.146
		ArtificiallyInspired	45.855	0.332 ± 0.131	15.92 ± 4.388	0.652 ± 0.156
		SIGMA	50.438	0.316 ± 0.121	17.787 ± 4.43	0.673 ± 0.152
		CoModGan [69]	108.929	0.516 ± 0.186	10.356 ± 3.024	0.502 ± 0.170
		LaMa [50]	95.278	0.529 ± 0.179	11.635 ± 3.500	0.345 ± 0.173
	EveryNLines	AIIA	7.405	0.109 ± 0.062	34.093 ± 4.719	0.931 ± 0.050
		HSSLAB	9.905	0.116 ± 0.061	33.578 ± 4.564	0.920 ± 0.053
		KwaiInpainting	9.558	0.135 ± 0.062	30.75 ± 4.727	0.914 ± 0.053
		ArtificiallyInspired	5.602	0.070 ± 0.049	36.178 ± 5.426	0.955 ± 0.037
		SIGMA	23.168	0.250 ± 0.098	28.503 ± 4.177	0.830 ± 0.088
		CoModGan [69]	118.959	0.575 ± 0.163	13.096 ± 3.173	0.400 ± 0.188
		LaMa [50]	146.771	0.731 ± 0.140	11.049 ± 3.090	0.098 ± 0.125
	Expand	AIIA	139.489	0.598 ± 0.072	12.811 ± 3.210	0.404 ± 0.158
		HSSLAB	112.939	0.578 ± 0.082	13.748 ± 3.345	0.429 ± 0.165
		KwaiInpainting	154.693	0.599 ± 0.076	13.359 ± 3.084	0.418 ± 0.159
		ArtificiallyInspired	69.536	0.624 ± 0.079	11.592 ± 2.731	0.357 ± 0.135
		SIGMA	102.994	0.605 ± 0.073	13.303 ± 2.939	0.390 ± 0.155
		CoModGan [69]	314.232	0.902 ± 0.078	6.488 ± 1.795	0.108 ± 0.061
		LaMa [50]	281.157	0.908 ± 0.101	9.134 ± 2.889	0.065 ± 0.052
	MediumStrokes	AIIA	28.116	0.132 ± 0.076	26.191 ± 6.142	0.869 ± 0.090
		HSSLAB	26.233	0.134 ± 0.070	26.149 ± 5.611	0.860 ± 0.089
		KwaiInpainting	48.812	0.155 ± 0.085	25.536 ± 5.879	0.877 ± 0.085
		ArtificiallyInspired	25.331	0.139 ± 0.079	24.860 ± 6.324	0.856 ± 0.097
		SIGMA	29.411	0.143 ± 0.080	25.602 ± 6.168	0.859 ± 0.095
		CoModGan [69]	170.570	0.291 ± 0.149	13.871 ± 5.134	0.746 ± 0.138
		LaMa [50]	173.157	0.337 ± 0.150	13.293 ± 4.573	0.519 ± 0.219
	NearestNeighbor	AIIA	25.149	0.283 ± 0.08	26.735 ± 4.683	0.729 ± 0.147
		HSSLAB	61.650	0.370 ± 0.156	25.414 ± 5.505	0.640 ± 0.197
		KwaiInpainting	47.438	0.397 ± 0.102	23.785 ± 4.051	0.652 ± 0.158
		ArtificiallyInspired	31.915	0.284 ± 0.115	28.126 ± 5.519	0.777 ± 0.139
		SIGMA	34.588	0.386 ± 0.077	22.878 ± 3.849	0.578 ± 0.144
		CoModGan [69]	233.938	0.612 ± 0.177	8.598 ± 1.719	0.128 ± 0.136
		LaMa [50]	186.051	0.588 ± 0.113	7.753 ± 3.364	0.046 ± 0.075
	ThickStrokes	AIIA	33.619	0.151 ± 0.082	23.887 ± 6.309	0.848 ± 0.095
		HSSLAB	31.389	0.155 ± 0.076	23.906 ± 5.940	0.839 ± 0.094
		KwaiInpainting	51.022	0.168 ± 0.086	22.993 ± 5.961	0.853 ± 0.092
		ArtificiallyInspired	28.795	0.157 ± 0.084	22.401 ± 6.511	0.831 ± 0.103
		SIGMA	30.851	0.157 ± 0.083	23.647 ± 6.276	0.837 ± 0.101
		CoModGan [69]	132.012	0.285 ± 0.139	13.453 ± 4.767	0.737 ± 0.134
		LaMa [50]	138.143	0.327 ± 0.136	13.150 ± 4.358	0.504 ± 0.211
	ThinStrokes	AIIA	19.28	0.117 ± 0.061	28.883 ± 5.021	0.893 ± 0.070
		HSSLAB	17.773	0.116 ± 0.053	28.862 ± 4.854	0.883 ± 0.071
		KwaiInpainting	36.549	0.149 ± 0.075	27.91 ± 4.805	0.897 ± 0.067
		ArtificiallyInspired	20.338	0.131 ± 0.065	27.517 ± 5.226	0.879 ± 0.079
		SIGMA	27.054	0.146 ± 0.071	27.325 ± 4.836	0.875 ± 0.078
		CoModGan [69]	218.699	0.409 ± 0.170	13.510 ± 4.182	0.716 ± 0.138
		LaMa [50]	218.117	0.443 ± 0.167	12.897 ± 3.772	0.499 ± 0.208

Table 13. Detailed Test Track 1 - WikiArt dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑
ImageNet	Completion	AIIA	53.329	0.314 ± 0.125	16.821 ± 4.050	0.687 ± 0.151
		HSSLAB	50.698	0.306 ± 0.122	16.943 ± 4.173	0.678 ± 0.151
		KwaiInpainting	66.676	0.331 ± 0.127	16.111 ± 3.877	0.683 ± 0.152
		ArtificiallyInspired	64.265	0.335 ± 0.131	14.528 ± 3.847	0.638 ± 0.159
		SIGMA	65.864	0.364 ± 0.118	16.946 ± 3.801	0.663 ± 0.155
		CoModGan [69]	68.808	0.492 ± 0.185	10.183 ± 3.247	0.510 ± 0.167
		LaMa [50]	60.203	0.533 ± 0.182	11.065 ± 3.246	0.269 ± 0.149
	EveryNLines	AIIA	4.358	0.091 ± 0.044	31.922 ± 4.940	0.930 ± 0.046
		HSSLAB	6.219	0.105 ± 0.056	30.62 ± 4.634	0.906 ± 0.056
		KwaiInpainting	7.734	0.132 ± 0.063	29.767 ± 4.684	0.913 ± 0.055
		ArtificiallyInspired	3.528	0.060 ± 0.037	33.457 ± 5.220	0.950 ± 0.036
		SIGMA	27.234	0.302 ± 0.102	25.918 ± 3.441	0.808 ± 0.098
		CoModGan [69]	85.176	0.537 ± 0.152	13.005 ± 2.707	0.420 ± 0.175
		LaMa [50]	113.321	0.700 ± 0.140	10.597 ± 2.512	0.104 ± 0.111
	Expand	AIIA	105.253	0.587 ± 0.082	12.958 ± 2.911	0.416 ± 0.166
		HSSLAB	91.267	0.571 ± 0.092	12.779 ± 3.181	0.423 ± 0.171
		KwaiInpainting	133.868	0.611 ± 0.085	11.859 ± 2.956	0.407 ± 0.160
		ArtificiallyInspired	119.22	0.621 ± 0.076	10.370 ± 2.340	0.333 ± 0.136
		SIGMA	135.302	0.634 ± 0.069	12.391 ± 2.423	0.380 ± 0.145
		CoModGan [69]	191.674	0.843 ± 0.100	7.657 ± 2.288	0.119 ± 0.073
		LaMa [50]	175.469	0.863 ± 0.121	9.103 ± 3.201	0.051 ± 0.046
	MediumStrokes	AIIA	6.898	0.049 ± 0.035	31.553 ± 7.784	0.948 ± 0.043
		HSSLAB	7.594	0.058 ± 0.040	29.857 ± 5.592	0.927 ± 0.050
		KwaiInpainting	11.599	0.065 ± 0.045	30.143 ± 8.024	0.947 ± 0.043
		ArtificiallyInspired	8.666	0.057 ± 0.040	30.151 ± 7.963	0.942 ± 0.048
		SIGMA	14.748	0.123 ± 0.054	28.136 ± 4.072	0.930 ± 0.051
		CoModGan [69]	61.314	0.160 ± 0.093	18.017 ± 5.878	0.881 ± 0.077
		LaMa [50]	65.721	0.233 ± 0.095	14.851 ± 3.603	0.474 ± 0.196
	NearestNeighbor	AIIA	21.180	0.263 ± 0.078	24.735 ± 4.443	0.723 ± 0.144
		HSSLAB	63.128	0.391 ± 0.172	23.173 ± 4.880	0.605 ± 0.196
		KwaiInpainting	68.085	0.420 ± 0.120	22.301 ± 3.913	0.638 ± 0.158
		ArtificiallyInspired	30.663	0.271 ± 0.109	25.729 ± 5.117	0.767 ± 0.144
		SIGMA	43.330	0.445 ± 0.078	20.566 ± 3.262	0.524 ± 0.137
		CoModGan [69]	190.993	0.612 ± 0.145	8.537 ± 1.455	0.118 ± 0.110
		LaMa [50]	158.894	0.572 ± 0.103	7.635 ± 2.709	0.050 ± 0.076
	ThickStrokes	AIIA	15.550	0.097 ± 0.066	26.507 ± 7.070	0.896 ± 0.080
		HSSLAB	14.551	0.100 ± 0.063	26.049 ± 5.748	0.878 ± 0.081
		KwaiInpainting	22.428	0.118 ± 0.077	25.181 ± 6.927	0.895 ± 0.080
		ArtificiallyInspired	17.582	0.108 ± 0.073	25.126 ± 7.259	0.884 ± 0.088
		SIGMA	23.115	0.170 ± 0.078	24.985 ± 4.831	0.877 ± 0.088
		CoModGan [69]	68.053	0.220 ± 0.134	15.157 ± 5.245	0.801 ± 0.127
		LaMa [50]	74.014	0.289 ± 0.128	13.325 ± 3.766	0.425 ± 0.193
	ThinStrokes	AIIA	5.534	0.050 ± 0.032	32.095 ± 5.683	0.950 ± 0.041
		HSSLAB	6.409	0.057 ± 0.037	30.703 ± 5.033	0.928 ± 0.048
		KwaiInpainting	9.090	0.067 ± 0.043	31.274 ± 5.369	0.949 ± 0.042
		ArtificiallyInspired	7.748	0.065 ± 0.049	30.593 ± 5.868	0.941 ± 0.048
		SIGMA	15.803	0.141 ± 0.061	28.242 ± 3.679	0.922 ± 0.055
		CoModGan [69]	110.980	0.283 ± 0.138	16.399 ± 4.132	0.832 ± 0.097
		LaMa [50]	113.163	0.337 ± 0.132	14.248 ± 3.335	0.447 ± 0.193

Table 14. Detailed Test Track 1 - ImageNet dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑	mIoU↑
FFHQ	Completion	MGTV	22.042	0.208 ± 0.082	17.024 ± 3.660	0.766 ± 0.099	0.955
		Baidu	18.479	0.187 ± 0.081	18.114 ± 3.847	0.774 ± 0.097	0.952
		HSSLAB	32.675	0.271 ± 0.124	18.412 ± 3.534	0.786 ± 0.097	0.707
		ArtificiallyInspired	19.545	0.209 ± 0.085	17.425 ± 3.446	0.756 ± 0.103	0.934
	EveryNLines	MGTV	1.785	0.016 ± 0.011	37.133 ± 1.999	0.970 ± 0.018	0.983
		Baidu	1.630	0.014 ± 0.008	38.261 ± 2.720	0.967 ± 0.016	0.985
		HSSLAB	5.615	0.031 ± 0.038	39.714 ± 4.572	0.959 ± 0.044	0.984
		ArtificiallyInspired	1.083	0.011 ± 0.009	41.779 ± 2.975	0.984 ± 0.009	0.987
	Expand	MGTV	54.141	0.397 ± 0.081	12.859 ± 2.646	0.567 ± 0.102	0.941
		Baidu	32.759	0.399 ± 0.080	12.880 ± 2.558	0.561 ± 0.097	0.928
		HSSLAB	101.753	0.513 ± 0.092	13.681 ± 2.321	0.594 ± 0.092	0.563
		ArtificiallyInspired	33.632	0.410 ± 0.085	13.318 ± 2.458	0.547 ± 0.096	0.898
	MediumStrokes	MGTV	13.758	0.086 ± 0.048	26.513 ± 5.686	0.894 ± 0.060	0.967
		Baidu	10.665	0.070 ± 0.042	26.944 ± 6.015	0.897 ± 0.059	0.971
		HSSLAB	12.837	0.075 ± 0.051	27.441 ± 6.805	0.921 ± 0.053	0.932
		ArtificiallyInspired	12.347	0.080 ± 0.046	26.504 ± 5.963	0.890 ± 0.062	0.961
	NearestNeighbor	MGTV	11.264	0.084 ± 0.040	31.285 ± 3.104	0.876 ± 0.054	0.955
		Baidu	7.992	0.065 ± 0.029	31.016 ± 2.964	0.863 ± 0.052	0.966
		HSSLAB	24.288	0.176 ± 0.176	31.066 ± 6.309	0.843 ± 0.140	0.926
		ArtificiallyInspired	14.360	0.110 ± 0.056	33.215 ± 3.485	0.908 ± 0.044	0.942
	ThickStrokes	MGTV	12.059	0.092 ± 0.057	24.997 ± 6.217	0.890 ± 0.066	0.969
		Baidu	10.142	0.078 ± 0.051	25.837 ± 6.374	0.894 ± 0.065	0.971
		HSSLAB	13.883	0.099 ± 0.066	25.191 ± 6.945	0.906 ± 0.062	0.896
		ArtificiallyInspired	11.611	0.090 ± 0.056	25.060 ± 6.420	0.886 ± 0.069	0.962
	ThinStrokes	MGTV	13.731	0.062 ± 0.034	30.093 ± 4.085	0.913 ± 0.048	0.963
		Baidu	9.153	0.051 ± 0.029	30.039 ± 4.285	0.914 ± 0.048	0.969
		HSSLAB	9.042	0.039 ± 0.041	32.668 ± 5.754	0.954 ± 0.044	0.966
		ArtificiallyInspired	11.096	0.062 ± 0.033	29.811 ± 4.102	0.906 ± 0.052	0.959

Table 15. Detailed Validation Track 2 - FFHQ dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑	mIoU↑
Places	Completion	MGTV	40.974	0.287 ± 0.129	16.847 ± 4.094	0.703 ± 0.144	0.592
		Baidu	41.102	0.290 ± 0.135	16.047 ± 4.252	0.687 ± 0.150	0.485
		HSSLAB	45.657	0.313 ± 0.152	17.962 ± 4.278	0.735 ± 0.138	0.427
		ArtificiallyInspired	39.444	0.283 ± 0.135	16.379 ± 4.275	0.678 ± 0.155	0.556
	EveryNLines	MGTV	3.689	0.029 ± 0.026	33.580 ± 3.998	0.958 ± 0.029	0.871
		Baidu	3.645	0.027 ± 0.023	33.445 ± 4.748	0.947 ± 0.037	0.892
		HSSLAB	3.999	0.027 ± 0.028	36.246 ± 4.628	0.964 ± 0.037	0.879
		ArtificiallyInspired	2.413	0.022 ± 0.024	36.504 ± 5.134	0.973 ± 0.024	0.903
	Expand	MGTV	75.698	0.537 ± 0.109	13.145 ± 2.791	0.443 ± 0.142	0.369
		Baidu	69.913	0.552 ± 0.102	11.828 ± 2.586	0.414 ± 0.141	0.217
		HSSLAB	93.050	0.590 ± 0.124	13.968 ± 2.870	0.499 ± 0.150	0.170
		ArtificiallyInspired	60.889	0.513 ± 0.115	12.546 ± 3.100	0.401 ± 0.154	0.353
	MediumStrokes	MGTV	24.126	0.113 ± 0.068	25.047 ± 6.055	0.874 ± 0.085	0.759
		Baidu	18.578	0.091 ± 0.058	24.885 ± 6.144	0.877 ± 0.083	0.789
		HSSLAB	17.306	0.075 ± 0.054	26.946 ± 7.173	0.910 ± 0.072	0.805
		ArtificiallyInspired	22.390	0.111 ± 0.067	24.330 ± 6.174	0.864 ± 0.090	0.749
	NearestNeighbor	MGTV	18.783	0.174 ± 0.096	26.544 ± 4.110	0.805 ± 0.113	0.681
		Baidu	15.424	0.127 ± 0.069	26.022 ± 4.517	0.776 ± 0.131	0.728
		HSSLAB	25.402	0.182 ± 0.186	29.040 ± 7.111	0.819 ± 0.182	0.550
		ArtificiallyInspired	18.713	0.165 ± 0.102	28.046 ± 4.987	0.841 ± 0.109	0.663
	ThickStrokes	MGTV	24.911	0.129 ± 0.077	23.489 ± 6.603	0.859 ± 0.092	0.739
		Baidu	21.766	0.116 ± 0.072	23.542 ± 6.598	0.865 ± 0.089	0.745
		HSSLAB	21.813	0.108 ± 0.073	24.926 ± 7.778	0.888 ± 0.082	0.735
		ArtificiallyInspired	24.390	0.128 ± 0.076	22.700 ± 6.843	0.846 ± 0.099	0.730
	ThinStrokes	MGTV	18.866	0.084 ± 0.045	27.526 ± 4.947	0.898 ± 0.065	0.762
		Baidu	12.920	0.065 ± 0.040	27.543 ± 5.075	0.902 ± 0.064	0.818
		HSSLAB	8.536	0.036 ± 0.029	31.088 ± 6.101	0.947 ± 0.053	0.874
		ArtificiallyInspired	18.226	0.093 ± 0.048	26.808 ± 4.883	0.885 ± 0.070	0.741

Table 16. Detailed Validation Track 2 - Places dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑	mIoU↑
FFHQ	Completion	MGTV	21.300	0.252 ± 0.090	17.094 ± 3.735	0.764 ± 0.096	0.953
		Baidu	18.105	0.232 ± 0.089	18.143 ± 4.135	0.771 ± 0.097	0.950
		HSSLAB	33.486	0.300 ± 0.121	17.128 ± 3.665	0.757 ± 0.101	0.690
		ArtificiallyInspired	19.179	0.260 ± 0.095	17.465 ± 3.751	0.753 ± 0.103	0.932
	EveryNLines	MGTV	1.725	0.050 ± 0.019	37.135 ± 2.018	0.970 ± 0.015	0.983
		Baidu	1.564	0.056 ± 0.026	38.265 ± 2.734	0.966 ± 0.016	0.985
		HSSLAB	6.055	0.087 ± 0.055	37.019 ± 4.068	0.958 ± 0.027	0.976
		ArtificiallyInspired	1.011	0.032 ± 0.019	41.778 ± 2.975	0.984 ± 0.009	0.987
	Expand	MGTV	51.917	0.460 ± 0.066	12.934 ± 2.715	0.566 ± 0.104	0.942
		Baidu	32.220	0.460 ± 0.074	13.145 ± 2.729	0.566 ± 0.098	0.928
		HSSLAB	103.304	0.536 ± 0.077	12.750 ± 2.227	0.563 ± 0.093	0.565
		ArtificiallyInspired	33.081	0.485 ± 0.073	13.347 ± 2.576	0.545 ± 0.098	0.895
	MediumStrokes	MGTV	13.327	0.124 ± 0.065	26.669 ± 5.778	0.893 ± 0.062	0.967
		Baidu	10.300	0.107 ± 0.057	27.047 ± 5.885	0.897 ± 0.060	0.970
		HSSLAB	17.545	0.137 ± 0.087	26.147 ± 5.730	0.884 ± 0.072	0.916
		ArtificiallyInspired	11.690	0.126 ± 0.066	26.724 ± 5.963	0.890 ± 0.064	0.961
	NearestNeighbor	MGTV	11.862	0.174 ± 0.058	31.205 ± 3.188	0.874 ± 0.057	0.954
		Baidu	8.401	0.161 ± 0.043	30.941 ± 3.039	0.861 ± 0.055	0.965
		HSSLAB	39.959	0.335 ± 0.172	29.002 ± 4.705	0.798 ± 0.110	0.876
		ArtificiallyInspired	14.966	0.178 ± 0.067	33.159 ± 3.556	0.907 ± 0.046	0.941
	ThickStrokes	MGTV	11.898	0.120 ± 0.067	25.164 ± 6.150	0.890 ± 0.067	0.970
		Baidu	10.237	0.106 ± 0.061	26.073 ± 6.192	0.895 ± 0.065	0.970
		HSSLAB	17.633	0.145 ± 0.086	24.150 ± 5.807	0.878 ± 0.072	0.884
		ArtificiallyInspired	11.525	0.123 ± 0.069	25.351 ± 6.321	0.887 ± 0.070	0.962
	ThinStrokes	MGTV	13.003	0.124 ± 0.060	30.185 ± 4.245	0.912 ± 0.049	0.964
		Baidu	8.541	0.100 ± 0.051	30.166 ± 4.413	0.913 ± 0.049	0.969
		HSSLAB	12.709	0.111 ± 0.074	30.113 ± 5.090	0.908 ± 0.059	0.952
		ArtificiallyInspired	10.203	0.133 ± 0.061	29.962 ± 4.279	0.905 ± 0.053	0.960

Table 17. Detailed Test Track 2 - FFHQ dataset

	Mask	Team	FID↓	LPIPS↓	PSNR↑	SSIM↑	mIoU↑
Places	Completion	MGTV	42.804	0.307 ± 0.119	16.716 ± 3.974	0.689 ± 0.144	0.583
		Baidu	43.597	0.31 ± 0.127	15.808 ± 3.991	0.672 ± 0.152	0.460
		HSSLAB	46.380	0.316 ± 0.130	16.921 ± 4.244	0.705 ± 0.144	0.420
		ArtificiallyInspired	40.911	0.317 ± 0.124	16.011 ± 4.123	0.661 ± 0.156	0.541
	EveryNLines	MGTV	3.581	0.061 ± 0.031	33.547 ± 3.927	0.957 ± 0.031	0.873
		Baidu	3.613	0.067 ± 0.038	33.442 ± 4.701	0.947 ± 0.041	0.890
		HSSLAB	4.772	0.086 ± 0.067	34.130 ± 4.544	0.956 ± 0.038	0.855
		ArtificiallyInspired	2.301	0.041 ± 0.031	36.474 ± 5.112	0.973 ± 0.025	0.906
	Expand	MGTV	73.709	0.546 ± 0.082	13.237 ± 2.773	0.448 ± 0.148	0.393
		Baidu	70.481	0.565 ± 0.089	11.954 ± 2.578	0.420 ± 0.145	0.229
		HSSLAB	87.303	0.572 ± 0.092	13.244 ± 2.782	0.483 ± 0.152	0.179
		ArtificiallyInspired	60.163	0.557 ± 0.087	12.697 ± 3.027	0.409 ± 0.158	0.372
	MediumStrokes	MGTV	22.100	0.132 ± 0.073	25.067 ± 5.820	0.876 ± 0.083	0.760
		Baidu	17.331	0.112 ± 0.066	24.923 ± 6.082	0.879 ± 0.082	0.791
		HSSLAB	16.746	0.108 ± 0.059	25.878 ± 5.307	0.888 ± 0.075	0.789
		ArtificiallyInspired	20.718	0.141 ± 0.078	24.317 ± 6.085	0.865 ± 0.088	0.744
	NearestNeighbor	MGTV	18.498	0.225 ± 0.080	26.710 ± 3.953	0.809 ± 0.110	0.673
		Baidu	15.086	0.209 ± 0.073	26.187 ± 4.350	0.781 ± 0.128	0.725
		HSSLAB	37.341	0.291 ± 0.183	26.998 ± 6.276	0.765 ± 0.184	0.465
		ArtificiallyInspired	18.421	0.203 ± 0.092	28.213 ± 4.825	0.845 ± 0.107	0.655
	ThickStrokes	MGTV	25.047	0.145 ± 0.079	23.113 ± 5.952	0.856 ± 0.092	0.743
		Baidu	21.081	0.131 ± 0.076	22.935 ± 6.050	0.859 ± 0.090	0.751
		HSSLAB	22.035	0.132 ± 0.073	23.740 ± 5.687	0.867 ± 0.086	0.724
		ArtificiallyInspired	23.802	0.153 ± 0.083	22.279 ± 6.054	0.843 ± 0.098	0.735
	ThinStrokes	MGTV	18.180	0.127 ± 0.062	27.775 ± 4.812	0.902 ± 0.064	0.765
		Baidu	12.653	0.094 ± 0.051	27.773 ± 5.034	0.906 ± 0.062	0.830
		HSSLAB	9.804	0.081 ± 0.040	29.503 ± 4.723	0.920 ± 0.054	0.858
		ArtificiallyInspired	18.322	0.148 ± 0.070	27.088 ± 4.826	0.890 ± 0.069	0.743

Table 18. Detailed Test Track 2 - Places dataset