INVESTIGATING ROBUSTNESS OF UNSUPERVISED STYLEGAN IMAGE RESTORATION

I. Key Contributions

The key points and problems addressed in this work:

- Unsupervised Image Restoration
- Method
- Ablation
- Conclusion

II. Unsupervised Image Restoration

We performed restoration at five degradation levels: XL (Extra Low), L (Low), M (Medium), S (Severe), and XS (Extra Severe). These levels combine various degradation parameters tailored to the restoration task. We used similar parameters as those proposed in RUSIR [1].

The data presented in Tables I and Table II summarize the experimental results of image restoration methods applied to the Drone dataset across various types of degradation. The evaluation of these results is based on three key performance metrics: LPIPS, LPIPS-vgg, and PFID, which are assessed at different levels of image degradation. The goal is to compare the performance of our proposed method (referred to as ”Ours”) with the existing RUSIR method on these metrics. Lower values for these metrics indicate better performance, signifying results that are closer to the original image.

The tables presented in Table III and Table IV report the quantitative performance results of image restoration methods applied to the FFHQ dataset, assessing different degradation types across multiple metrics: LPIPS, LPIPS-vgg, and PFID. These results compare the proposed method (Ours) with the existing RUSIR method, with lower values of these metrics indicating better restoration quality.

III. Method

We explore StyleGAN based unsupervised image restoration by integration of various IQA-based loss functions into the StyleGAN image restoration pipeline, exploring their effects on the final restored image quality.

After generating synthetic images using StyleGAN2-ADA, we evaluate their quality through several metrics: Inception Score (IS) [2], Frechet Inception Distance (FID) [3], and Kernel Inception Distance (KID) [4]. These metrics provide a quantitative assessment of the generated images, measuring their realism and alignment with the distribution of real images. The Inception Score evaluates the clarity and diversity of the generated images, while FID and KID assess the similarity between the synthetic and real image distributions, with FID focusing on the distance between feature distributions and KID measuring discrepancies in feature distributions using kernel.

Degradation Type	LPIPS ↓		LPIPS-vgg ↓		PFID ↓	
	RUSIR [1]	*Ours*	*RUSIR [1]*	*Ours*	*RUSIR [1]*	*Ours*
Upsampling						
XS	0.747	0.649	0.745	0.709	115.7	67.3
S	0.814	0.737	0.755	0.719	122.2	58.3
M	0.818	0.729	0.760	0.722	123.4	95.5
L	0.828	0.756	0.769	0.735	152.6	133.2
XL	0.855	0.807	0.784	0.757	182.6	173.1
Denoising						
XS	0.759	0.647	0.746	0.708	122.5	61.2
S	0.769	0.653	0.747	0.709	124.6	59.1
M	0.788	0.666	0.748	0.706	122.0	66.6
L	0.800	0.683	0.749	0.707	132.1	81.0
XL	0.806	0.704	0.750	0.720	136.0	117.2
Deartifacting						
XS	0.758	0.626	0.738	0.681	140.5	74.2
S	0.765	0.630	0.740	0.687	145.1	74.8
M	0.768	0.639	0.746	0.690	145.8	80.9
L	0.778	0.660	0.750	0.699	142.7	93.0
XL	0.818	0.688	0.766	0.714	158.4	121.1
Inpainting						
XS	0.745	0.649	0.725	0.676	128.5	81.5
S	0.735	0.648	0.723	0.677	153.3	86.1
M	0.740	0.650	0.728	0.683	159.9	89.6
L	0.738	0.656	0.730	0.690	170.8	99.0
XL	0.747	0.658	0.736	0.697	164.9	112.9

TABLE I: Experimental results on the Drone dataset, for single degradation. The lowest values in each column indicate the best results.

A. Latent Space Extension and Optimization

Image restoration begins with StyleGAN inversion, where we seek a latent code $w \in W$ such that the generated image $x_{clean} = G(w)$ closely matches the degraded target image $y = f(y_{clean})$. To improve fidelity, the latent space is extended to $W+$ and further to $W++$, providing more degrees of freedom to better align the generated image with the target.

B. IQA-Based and LPIPS Loss Functions

We aim to improve image restoration quality by incorporating multiple Image Quality Assessment (IQA)-based loss functions to guide the StyleGAN inversion to capture diverse perceptual and structural aspects of image quality, leading to a more robust and nuanced restoration process.

LPIPS Loss: The LPIPS score measures perceptual similarity between the activations of two image patches for extracted features from a pre-trained network. We use the LPIPS score to define a loss function as follows.

Traditional loss functions like MSE and L1 distance, which compare images at the pixel level, often fail to capture the perceptual differences between images. LPIPS is more tolerant of small spatial

Degradation Type	LPIPS ↓		LPIPS-vgg ↓		PFID ↓	
	RUSIR [1]	*Ours*	*RUSIR [1]*	*Ours*	*RUSIR [1]*	*Ours*
2 Degradation						
NA	0.795	0.687	0.751	0.714	117.3	71.3
AP	0.732	0.655	0.738	0.697	164.6	108.9
UA	0.890	0.840	0.793	0.785	198.2	175.7
NP	0.779	0.684	0.747	0.714	155.7	98.6
UN	0.843	0.811	0.778	0.751	167.1	157.3
Degradation Type	**LPIPS ↓**		**LPIPS-vgg ↓**		**PFID ↓**	
	RUSIR [1]	*Ours*	*RUSIR [1]*	*Ours*	*RUSIR [1]*	*Ours*
3 degradation						
UNP	0.816	0.789	0.768	0.749	159.5	155.9
UPA	0.833	0.838	0.786	0.784	150.4	186.1
UNA	0.888	0.885	0.808	0.787	215.5	185.6
NAP	0.790	0.702	0.750	0.718	140.6	101.1
Degradation Type	**LPIPS ↓**		**LPIPS-vgg ↓**		**PFID ↓**	
	RUSIR [1]	*Ours*	*RUSIR [1]*	*Ours*	*RUSIR [1]*	*Ours*
4 degradation						
UNAP	0.840	0.859	0.796	0.782	174.2	231.3

TABLE II: Quantitative results for multiple degradation in the Drone dataset.

misalignments between images, as it focuses on overall perceptual similarity at feature space rather than exact pixel matches.

$$\ell_{\text{LPIPS}}(x, y) = \sum_i w_i \cdot \|\phi_i(x) - \phi_i(y)\|_2^2 \tag{1}$$

Here, x is the restored image and y is the reference image, $\phi_i(x)$ and $\phi_i(y)$ denote the feature maps of the two images at layer i of a pre-trained neural network, capturing the perceptual details at different levels. The term w_i represents learned weights that scale the contribution of each feature layer.

IV. Ablation

While our approach builds on the RUSIR framework, the primary innovation lies in the combination of multiple loss functions to improve robustness. improve restoration performance by leveraging complementary features such as perceptual similarity (LPIPS), structural integrity (MS-SSIM)(see Table V), Gradient(see Table VII) and consistency(see Table VI). This systematic integration of loss functions in StyleGAN-specific context offers new insights into the application of multi-loss functions in generative models, which, to the best of our knowledge, has not been explored in such detail in unsupervised image restoration tasks(see Fig. 1).

V. Conclusion

We have shown that integrating composite loss functions in the StyleGAN inversion significantly enhances performance to improve the robustness of unsupervised image restoration. We have shown through experiments that the proposed approach achieves high-fidelity and realistic image restoration while adapting to the specific nature of the degradation. In future, we propose improving the restoration quality in the presence of multiple degradations in the input

Degradation Type	LPIPS		LPIPS-vgg		PFID	
	RUSIR [1]	*Ours*	*RUSIR [1]*	*Ours*	*RUSIR [1]*	*Ours*
Upsampling						
XS	0.215	0.146	0.339	0.284	48.6	22.7
S	0.255	0.202	0.378	0.337	49.7	26.3
M	0.313	0.289	0.414	0.399	61.6	47.2
L	0.335	0.341	0.428	0.435	83.7	68.1
XL	0.353	0.365	0.439	0.450	95.2	93.6
Denoising						
XS	0.246	0.199	0.373	0.348	55.5	39.4
S	0.259	0.209	0.383	0.355	59.5	45.8
M	0.278	0.241	0.397	0.378	63.5	51.3
L	0.306	0.278	0.414	0.400	70.4	58.1
XL	0.338	0.386	0.430	0.459	83.5	79.0
Deartifacting						
XS	0.264	0.188	0.380	0.337	56.3	33.9
S	0.272	0.201	0.385	0.350	60.8	33.9
M	0.282	0.212	0.394	0.361	65.2	39.5
L	0.297	0.242	0.406	0.387	68.1	46.4
XL	0.343	0.304	0.438	0.432	81.5	60.1
Inpainting						
XS	0.203	0.128	0.300	0.231	51.9	24.3
S	0.212	0.152	0.314	0.262	63.1	36.9
M	0.225	0.173	0.329	0.286	68.9	42.0
L	0.241	0.193	0.343	0.307	71.6	46.9
XL	0.256	0.213	0.357	0.325	77.0	59.3

TABLE III: Quantitative Results for samples of FFHQ-dataset, for single degradation.

References

[1] Y. Poirier-Ginter and J.-F. Lalonde, "Robust unsupervised stylegan image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 292–22 301.

[2] S. Barratt and R. Sharma, "A note on the inception score," *arXiv preprint arXiv:1801.01973*, 2018.

[3] A. Obukhov and M. Krasnyanskiy, "Quality assessment method for gan based on modified metrics inception score and fréchet inception distance," in *Software Engineering Perspectives in Intelligent Systems: Proceedings of 4th Computational Methods in Systems and Software 2020, Vol. 1 4*. Springer, 2020, pp. 102–114.

[4] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying mmd gans," *arXiv preprint arXiv:1801.01401*, 2018.

Degradation Type	LPIPS		LPIPS-vgg		PFID	
	RUSIR [1]	*Ours*	*RUSIR [1]*	*Ours*	*RUSIR [1]*	*Ours*
2 degradation						
NA	0.301	0.296	0.417	0.422	70.358	57.3
AP	0.302	0.252	0.408	0.386	83.021	55.9
UA	0.375	0.431	0.455	0.492	102.8	96.0
NP	0.309	0.277	0.412	0.395	84.865	63.9
UN	0.377	0.460	0.455	0.513	107.0	120.4
Degradation Type	**LPIPS**		**LPIPS-vgg**		**PFID**	
	RUSIR [1]	*Ours*	*RUSIR [1]*	*Ours*	*RUSIR [1]*	*Ours*
3 degradation						
UNP	0.380	0.459	0.458	0.502	107.683	115.2
UPA	0.387	0.441	0.459	0.498	114.313	107.9
UNA	0.404	0.485	0.475	0.534	116.477	122.3
NAP	0.328	0.302	0.428	0.421	85.779	68.0
Degradation Type	**LPIPS**		**LPIPS-vgg**		**PFID**	
	RUSIR [1]	*Ours*	*RUSIR [1]*	*Ours*	*RUSIR [1]*	*Ours*
4 degradation						
UNAP	0.404	0.475	0.475	0.519	121.924	110.3

TABLE IV: Quantitative results for multiple degradation in the samples of FFHQ-dataset.

Task	SSIM	PSNR	LPIPS	FID
Upsampling	0.765	23.592	0.357	104.832
Denoising	0.424	12.311	0.693	149.733
Deartifacting	0.799	21.304	0.419	133.710
Inpainting	0.801	21.268	0.441	137.441

TABLE V: Comparison of different tasks based on SSIM loss for L degradation

Task	SSIM	PSNR	LPIPS	FID
Upsampling	0.581	11.205	0.618	250.330
Denoising	0.010	5.252	0.917	600.371
Deartifacting	0.568	10.350	0.628	182.790
Inpainting	0.581	11.205	0.618	250.330

TABLE VI: Comparison of different tasks based on Consistency loss for L degradation

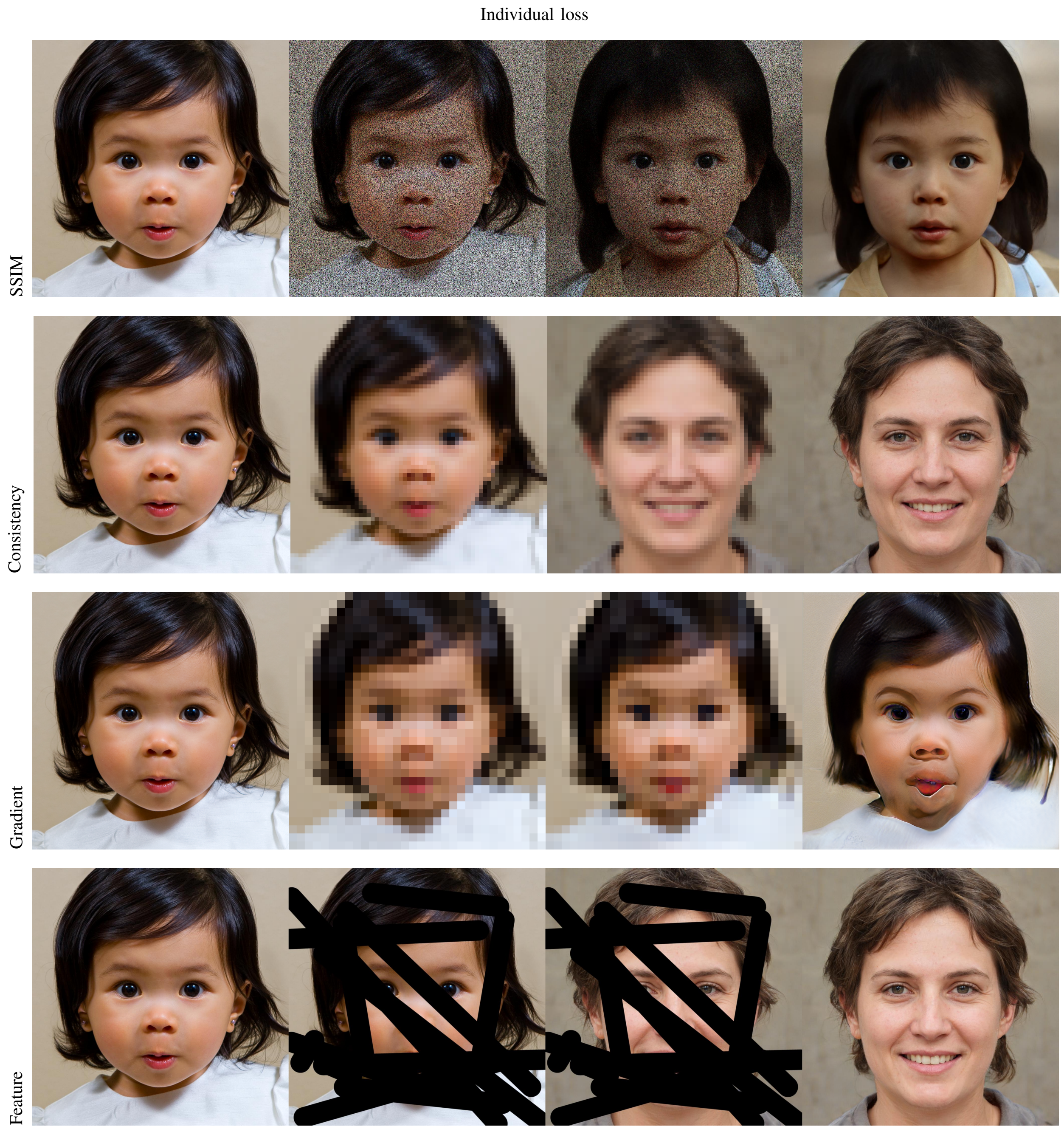

Fig. 1: This figure demonstrates the process for each element, where the first image shows the input, followed by the degradation of the image according to the specified binomial and Poisson distributions. Subsequently, leveraging prior information, the model attempts to reconstruct the image to its original form, showcasing the effectiveness of our method.

Task	**SSIM**	**PSNR**	**LPIPS**	**FID**
Upsampling	0.571	11.081	0.619	210.182
Denoising	0.579	11.997	0.583	199.621
Deartifacting	0.569	11.167	0.614	213.464
Inpainting	0.587	12.005	0.582	206.164

TABLE VII: Comparison of different tasks based on Gradient loss for L degradation