Supplementary Material For BIGPrior: Towards Decoupling Learned Prior Hallucination and Data Fidelity in Image Restoration

Majed El Helou, *Member, IEEE,* Sabine Süsstrunk, *Fellow, IEEE.*

***Abstract*—We present further experiments on the image colorization, inpainting, randomized-masking inpainting, and AWGN removal restoration tasks. The extended results are specifically on test data taken from each generative network's training distribution. This experimental configuration enables us to assess potential overfitting in the generative pre-training. In addition, it enables us to evaluate to which degree data fidelity could improve generative prior results even on data seen by that generative network during its training. Our results on these experiments support two points. First, the generative network has good distribution modeling since its results on the test images are not significantly worse than on the training data as would be the case with overfitting. Second, our framework consistently improves the results on the various image sets, even those extracted from the generative network's training distribution, thus underlining the importance of data fidelity.**

I. Experiments and Discussion

Each of the presented results follows from the same experimental setup as the one presented in the main manuscript. The only difference is that we additionally test the mGAN prior method [2] on its generative network's training data, and similarly with our framework. We refer to this experimental testing setup as **I**n **G**enerative **D**istribution (IGD) in what follows. The image colorization results are reported in Table I. We report the results of central image inpainting in Table II, and of randomized-masking inpainting in Table III. The results of AWGN removal are lastly reported in Table IV.

We can note with the colorization experiment that the mGAN results increase by 2.05AuC on the IGD Bedroom set, and actually decrease by 0.56AuC on the IGD Church set. This strongly underlines that there was no overfitting in the PGGAN pre-trained on the Church dataset. We discuss the Bedroom dataset in the last paragraph of this section. As for our method, we reach +2.58AuC on IGD data from the Bedroom set relative to our regular results, surpassing the IGD mGAN results by +0.28AuC, and while the results decrease on the IGD Church data, they still surpass the mGAN results by +0.84AuC. This trend is consolidated across the various other experiments. For central image inpainting, our results improve by +0.17PSNR on the IGD data, while still surpassing the mGAN results by +4.2PSNR, +0.32SSIM, and -0.16LPIPS. If we consider the Conference dataset, the mGAN results improve for IGD data by +0.76PSNR on randomized-masking inpainting while our results still surpass them by +1.97PSNR. Similarly, for AWGN removal they improve by +0.9PSNR while ours are still +3.25PSNR higher. The general observation we make across our experiments is that the generative network inversion results are not better by a large margin on the data taken from the generative network's training distribution relative to the ones on the mutually exclusive test sets. Additionally, we observe that our framework, capable of exploiting data fidelity, consistently improves the results even when the prior inversion is carried out on IGD images.

Nevertheless, we notice certain cases (for instance in randomized-inpainting on the Bedroom dataset, and in AWGN removal on the Bedroom dataset) where the prior results improve but our method's results do not, on the IGD data. These results suggest that the pre-trained generative network suffered to a certain degree from overfitting in its learning on the Bedroom training set. This is further supported by the +2.05AuC obtained in the colorization experiment on the IGD Bedroom set. Due to this overfitting, it is better able to reconstruct the part of the restoration output that would normally be covered by the data-fidelity component on the IGD test data. This is witnessed, for instance, in the low noise level data, or the non-masked areas in the randomized-inpainting experiments. However, as this is already covered by our data-fidelity term, it does not contribute to improving the quality of the final results. For that, an improvement of the quality of the prior hallucination is needed, or, in other words, a better distribution modeling of the generative network rather than an overfitting to its training dataset.

References

[1] David Bau, Hendrik Strobelt, William Peebles, Bolei Zhou, Jun-Yan Zhu, Antonio Torralba, et al. Semantic photo manipulation with a generative image prior. *arXiv preprint arXiv:2005.07727*, 2020. 2
[2] Jinjin Gu, Yujun Shen, and Bolei Zhou. Image processing using multi-code GAN prior. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3012–3021, 2020. 1, 2
[3] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Computer Vision and Pattern Recognition (CVPR)*, pages 9446–9454, 2018. 2
[4] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Computer Vision and Pattern Recognition (CVPR)*, pages 5505–5514, 2018. 2
[5] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *International Conference on Computer Vision (ICCV)*, pages 4471–4480, 2019. 2
[6] Richard Zhang, Phillip Isola, and Alexei Efros. Colorful image colorization. In *European Conference on Computer Vision (ECCV)*, pages 649–666, 2016. 2

Method	Bedroom set AuC [6] ↑	Church set AuC [6] ↑
Colorful colorization [6]	88.55	89.13
Deep image prior [3]	84.33	83.31
Feature map opt. [1]	85.41	86.10
mGAN prior [2]	88.52	89.69
Ours	**89.27**	**90.64**
(IGD) mGAN prior [2]	90.57	89.13
(IGD) Ours	**91.85**	**89.97**

TABLE I

QUANTITATIVE AUC (%) RESULTS FOR IMAGE COLORIZATION ON THE BEDROOM AND CHURCH TEST SETS. THE HIGHER THE VALUE, THE LOWER IS THE CUMULATIVE COLORIZATION ERROR CURVE. THE LAST TWO ROWS SHOW THE RESULTS WITH DATA TAKEN FROM THE TRAINING SET OF THE GENERATIVE NETWORK, TO ANALYZE THE EFFECT OF DISTRIBUTION SHIFT ON THE PRIOR. THE BEST RESULTS FOR THE REGULAR TEST SETS AND FOR THE IGD TEST SETS ARE EACH HIGHLIGHTED IN BOLD, AND WE UNDERLINE THE SECOND BEST RESULTS ON THE REGULAR TEST SETS.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
DeepFill v2 [4], [5]	**26.56**	**0.9555**	**0.0191**
Feature map opt. [1]	14.75	0.4563	-
Deep image prior [3]	17.92	0.4327	-
mGAN prior [2]	20.55	0.5823	0.2070
Ours	25.32	0.9240	0.0376
(IGD) mGAN prior [2]	21.29	0.6013	0.1979
(IGD) Ours	**25.49**	**0.9261**	**0.0372**

TABLE II

QUANTITATIVE PSNR (dB), SSIM, AND LPIPS RESULTS FOR CENTRAL IMAGE INPAINTING. WE MASK OUT A 64×64 PATCH FROM THE CENTER OF EACH INPUT IMAGE. THE TASK-SPECIFIC STATE-OF-THE-ART METHOD IS HIGHLIGHTED WITH BACKGROUND SHADED IN GRAY. THE LAST TWO ROWS SHOW THE RESULTS WITH DATA TAKEN FROM THE TRAINING SET OF THE GENERATIVE NETWORK, TO ANALYZE THE EFFECT OF DISTRIBUTION SHIFT ON THE PRIOR. THE BEST RESULTS FOR THE REGULAR TEST SET AND FOR THE IGD TEST SET ARE EACH HIGHLIGHTED IN BOLD, AND WE UNDERLINE THE SECOND BEST RESULTS ON THE REGULAR TEST SETS.

Test	Method	PSNR ↑	SSIM ↑	LPIPS ↓
Bed.	mGAN prior [2]	20.34	0.5902	0.2134
Bed.	Ours	**23.22**	**0.8598**	**0.0775**
Chu.	mGAN prior [2]	19.33	0.5359	0.2235
Chu.	Ours	**21.94**	**0.8509**	**0.0855**
Conf.	mGAN prior [2]	19.38	0.5641	0.2062
Conf.	Ours	**22.20**	**0.8318**	**0.0785**
	In Generative Distribution (IGD)			
Bed.	mGAN prior [2]	20.84	0.6056	0.2073
Bed.	Ours	**22.19**	**0.7921**	**0.1044**
Chu.	mGAN prior [2]	19.60	0.5395	0.2159
Chu.	Ours	**22.12**	**0.8251**	**0.0904**
Conf.	mGAN prior [2]	20.14	0.5808	0.2074
Conf.	Ours	**22.11**	**0.8505**	**0.0843**

TABLE III

QUANTITATIVE PSNR (dB), SSIM, AND LPIPS RESULTS FOR RANDOMIZED-MASKING INPAINTING ON THE BEDROOM, CHURCH (OUTDOOR), AND CONFERENCE TEST SETS. THE RANDOMIZED MASKING INCREASES THE DIFFICULTY OF PREDICTING OUR ϕ MAPS. WE COMPARE THE PRIOR-BASED RESULTS TO OURS, TO ANALYZE THE EFFECT OF MASK RANDOMIZATION ON THE PERFORMANCE OUR ϕ PREDICTION COMPARED TO THE CENTRAL INPAINTING TASK. THE BOTTOM HALF SHOWS THE RESULTS WITH DATA TAKEN FROM THE TRAINING SET OF THE GENERATIVE NETWORK, TO ANALYZE THE EFFECT OF DISTRIBUTION SHIFT ON THE PRIOR.

Test	Method	PSNR ↑	SSIM ↑	LPIPS ↓
Bed.	mGAN prior [2]	22.72	0.6257	0.1978
Bed.	Ours	**26.80**	**0.7279**	**0.0998**
Chu.	mGAN prior [2]	21.12	0.5643	0.2065
Chu.	Ours	**23.38**	**0.5959**	**0.1435**
Conf.	mGAN prior [2]	21.49	0.5962	0.1968
Conf.	Ours	**24.70**	**0.6578**	**0.1192**
	In Generative Distribution (IGD)			
Bed.	mGAN prior [2]	23.29	0.6387	0.1930
Bed.	Ours	**26.16**	**0.6654**	**0.1257**
Chu.	mGAN prior [2]	21.56	0.5657	0.2020
Chu.	Ours	**24.63**	**0.6687**	**0.1171**
Conf.	mGAN prior [2]	22.39	0.6175	0.1884
Conf.	Ours	**25.64**	**0.6472**	**0.1242**

TABLE IV

PSNR (dB), SSIM, AND LPIPS RESULTS FOR AWGN REMOVAL ON THE BEDROOM, CHURCH, AND CONFERENCE SETS. THE NOISE FOLLOWS A GAUSSIAN DISTRIBUTION WITH STANDARD DEVIATION SAMPLED UNIFORMLY AT RANDOM FROM [5,50] PER IMAGE. THE BOTTOM HALF SHOWS THE RESULTS WITH DATA TAKEN FROM THE TRAINING SET OF THE GENERATIVE NETWORK, TO ANALYZE THE EFFECT OF DISTRIBUTION SHIFT ON THE PRIOR.