Supplementary Material
ISNAS-DIP: Image-Specific Neural Architecture Search for Deep Image Prior

Metin Ersin Arican[1]*, Ozgur Kara[1]*, Gustav Bredell[2] , Ender Konukoglu[2]
[1]Department of Electrical and Electronics Engineering, Bogazici University, Istanbul, Turkey
[2] Department of Information Technology and Electrical Engineering, ETH-Zurich, Zurich, Switzerland
[1]{metin.arican, ozgur.kara}@boun.edu.tr
[2]{gustav.bredell, ender.konukoglu}@vision.ee.ethz.ch

Overview

This supplementary file provides details and results in addition to the main manuscript. Furthermore, codes and dataset are available at https://github.com/ozgurkara99/ISNAS-DIP.

1 Deep Image Prior

In deep image prior, the network f_θ is enforced to map the noise z to the uncorrupted version of the corrupted image x_0 by minimizing the following objective functions with respect to θ depending on the task of interest, and stopping the optimization at a pre-determined iteration point to prevent overfitting. We follow DIP [1] and optimize the following objectives:

$$\mathcal{L}_{denoising}(\theta) = ||f_\theta(z) - x_0||^2 \tag{1}$$

$$\mathcal{L}_{inpainting}(\theta) = ||(f_\theta(z) - x_0) \otimes M||^2 \tag{2}$$

$$\mathcal{L}_{superresolution}(\theta) = ||D(f_\theta(z)) - x_0||^2 \tag{3}$$

where $D(\cdot)$ denotes the downsampling operator, $\otimes$ denotes pixel-wise multiplication, and M denotes the mask for inpainting.

2 Exponential Averaging

Inspired by DIP [1], we take the average of the reconstructed predictions x_t as our final image. Unlike DIP [1] and NAS-DIP [2], we did the averaging not only for image denoising but also for other tasks as well since we saw that it substantially improves the quality in all tasks. Note that we used exponential averaging for all approaches, i.e. DIP [1], NAS-DIP [2] and ISNAS-DIP. It is formulated as follows:

$$x^* = \gamma^{(T-1)} \cdot x_1 + \sum_{t=2}^{T} x_t \cdot \gamma^{T-t} \cdot (1-\gamma) \tag{4}$$

where x^* denotes the final result, T denotes the total iteration number, x_t denotes the restored output at t^{th} iteration, and γ is selected to be 0.99.

*equal contribution

3 Architecture Selection

We plot the normalized[1] PSNR increase scores for each model in *NAS Dataset for DIP* as shown in Fig. 1. Each subplot represents normalized PSNR increase scores for different image pairs. Models are sorted according to their normalized PSNR increase scores measured for the reference image (depicted with blue line) and evaluated on the target image (depicted with orange line). We select "chest" as our reference image since it belongs to a different domain as opposed to other images.

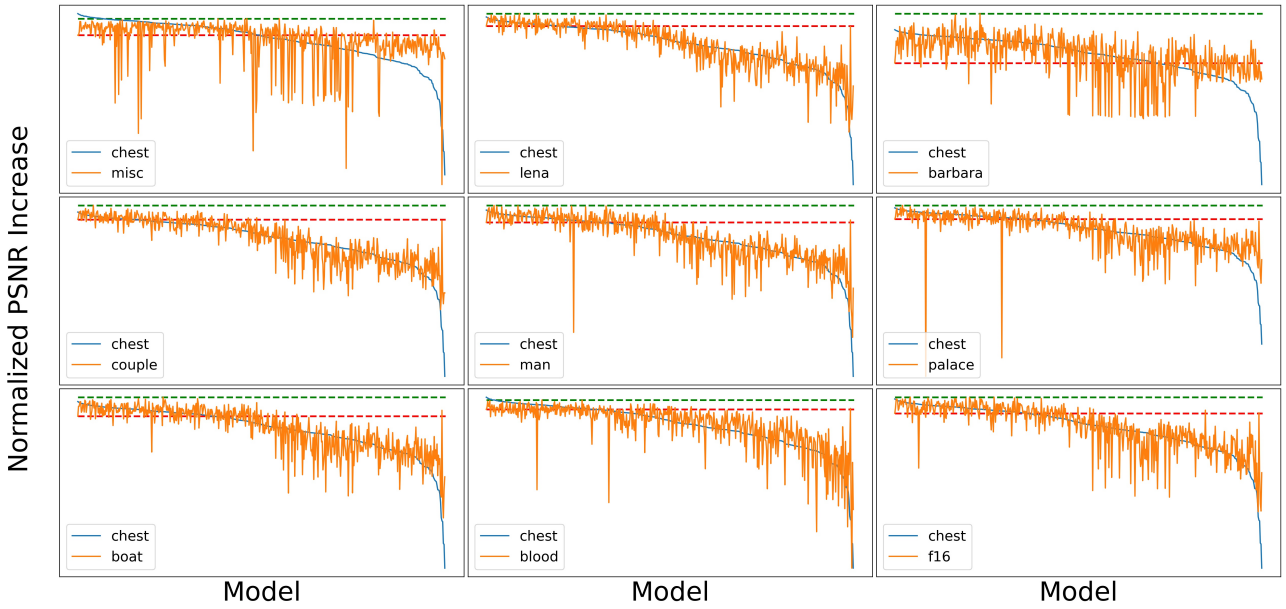


Figure 1: Normalized PSNR increase scores for each model in *NAS Dataset*. Each subplot corresponds to the scores of models that are sorted in descending order according to their scores on the "reference image" (blue line) and evaluated on the "target image" (orange line). Red dashed line depicts the normalized PSNR increase score of the best model of reference image evaluated on target image. Green dashed line shows the normalized PSNR increase score of the best model of the target image

As can be seen in Fig. 1, although the trend of the orange line is similar to that of reference image apparently; same models tend to perform differently. Furthermore, the worst models for the reference image may perform better on the target image.

The gap between green and red dashed line show how much the performance differs between best model for reference image and best model of the target image both evaluated on target image. For instance, we observe a large gap between the dashed lines for chest and barbara images due to the difference in domains that each belongs to.

4 Correlation between optimal stopping point

We observed that different models within the same search space have different optimal stopping points as shown in Fig. 2. We questioned if the optimal stopping point is an architectural property or not. In light of these questions, we report the correlations between our metrics and optimal stopping points as represented in Table 1. Surprisingly, there is a strong correlation between our metrics and the optimal stopping point. Our work will be extended by further investigation of the detection of optimal stopping point utilizing the proven strong correlation.

[1] We normalized the data by subtracting mean and dividing the deviation as follows: $x_{normalized} = \frac{x - mean}{std}$

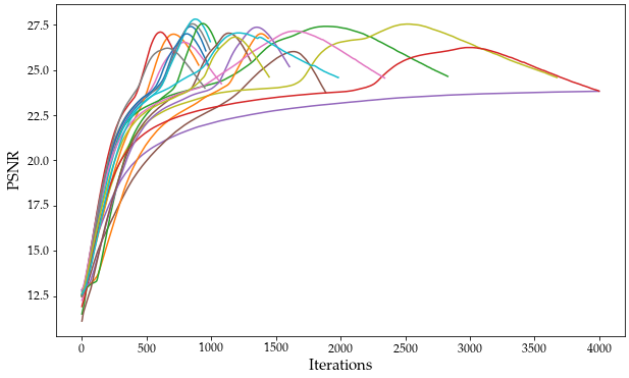


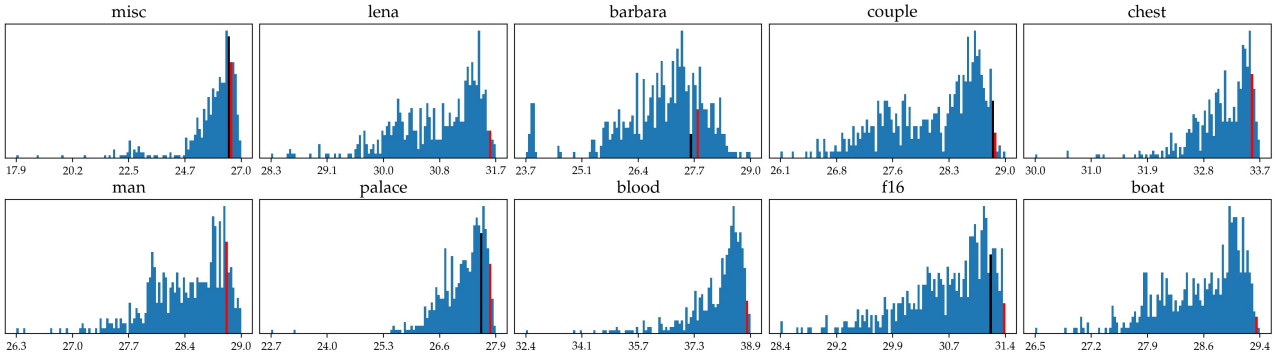
Figure 2: PSNR curve of different models optimized for the "barbara" image

Images	PSD DB MSE		PSD DB STRIP MSE		PSD STRIP HIST EMD		PSD 99 PER BW
	N	GT	N	GT	N	GT	N/GT
misc	0.24	-0.06	-0.45	-0.45	-0.44	-0.44	-0.39
lena	0.30	-0.40	-0.41	-0.41	-0.43	-0.43	-0.38
barbara	0.33	-0.43	-0.46	-0.46	-0.47	-0.47	-0.41
couple	0.28	-0.39	-0.40	-0.40	-0.43	-0.44	-0.36
chest	0.32	-0.36	-0.19	0.15	-0.40	-0.40	-0.34
man	0.28	-0.40	-0.41	-0.42	-0.44	-0.44	-0.37
palace	0.27	-0.35	-0.36	-0.37	-0.40	-0.40	-0.34
blood	0.20	-0.22	0.20	0.17	-0.06	0.04	-0.18
f16	0.28	-0.41	-0.42	-0.42	-0.43	-0.43	-0.38
boat	0.30	-0.41	-0.40	-0.40	-0.43	-0.43	-0.37

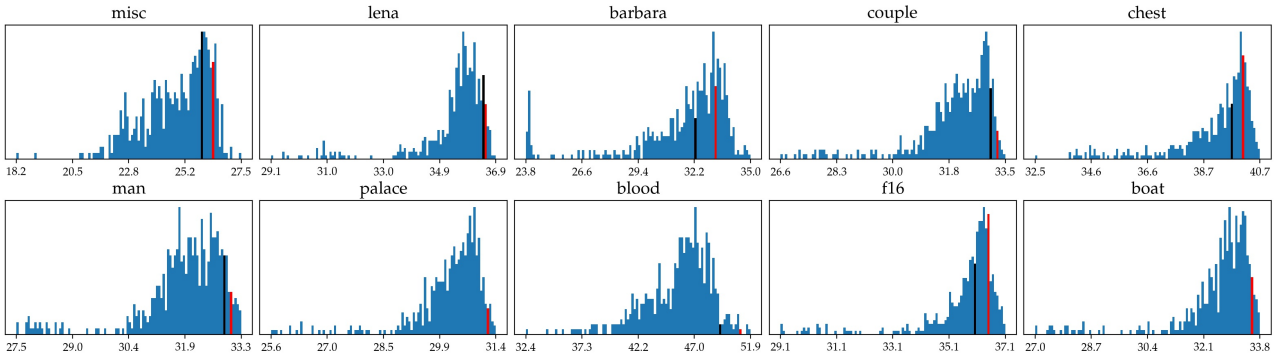
Table 1: Kendall correlation values between the optimal stopping points and corresponding metrics for image denoising task. MSE metrics are calculated either with the noisy image (denoted as N) or with the ground truth image (denoted as GT).

5 Histograms

Fig. 3 shows the distribution of the final PSNR scores of the 522 models for both image denoising (a) and image inpainting (b). Note that the best-performing model found with PSD DB Strip MSE metric is shown with red for N=15 and black for N=5 case.



(a) Denoising

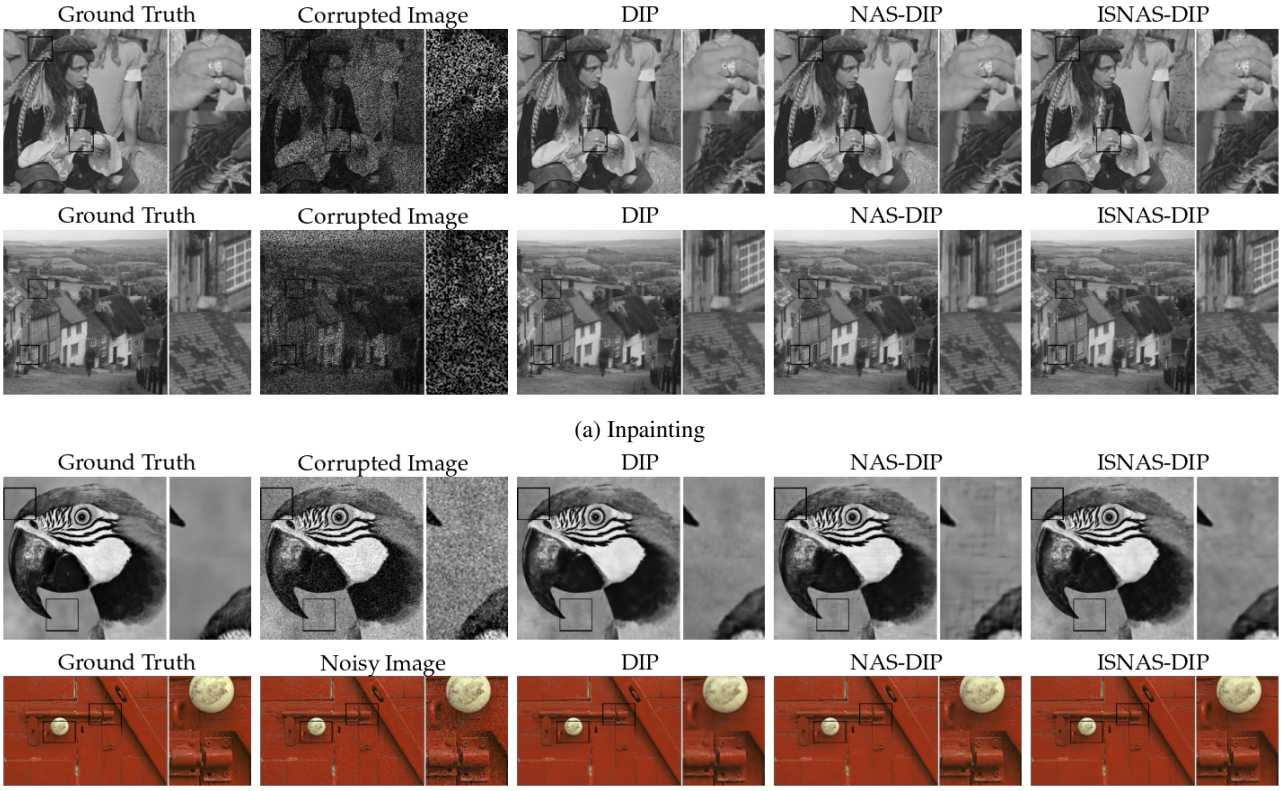


(b) Inpainting

Figure 3: Histograms of the final PSNR scores of models for all images in *the NAS Dataset for DIP* for image denoising and image inpainting . Non-blue colored bars represent the final PSNR scores of the best models included in top-N models found by PSD DB Strip MSE metric. The bin depicted with red denotes the best model among N=15 models, the black bin denotes the best one among N=5 models.

6 Qualitative Analysis

We show 2 more qualitative examples for denoising and inpainting tasks in Fig. 4.



(a) Inpainting

(b) Denoising

Figure 4: Qualitative examples of image inpainting (a), and image denoising (b) tasks. Each column shows the images of ground truth, corrupted image, reconstructed images of DIP [1], NAS-DIP [2], and ISNAS-DIP, respectively.

7 Super-resolution Results

Fig. 5 shows the performance of NAS-DIP [2] and our method on a super-resolution image with $\times 2$ scale. It is evident in the example that our method is not good at producing finer details since it tends to select models having more low-pass characteristics. Hence, it is not able to outperform previous approaches, indicating the need for a different metric for super-resolution.

(a) Ground Truth (b) NAS-DIP [2] (c) ISNAS-DIP

Figure 5: A qualitative example for super-resolution $\times 2$

8 Histogram Analysis with Random Selection

In Fig. 6, we additionally provide the resulting PSNR of the best performing models among a random selected 15 models 10 times, which is shown with yellow bin. In other words, at each time, we select 15 different models randomly, and record the best performing model's PSNR among these, and repeat this procedure 10 times to take the average. Whereas this is similar to the PSNR of the best model selected by the PSD DB Strip MSE metric it should be noted that it is not a deterministic selection process in contrast to our proposed method and can thus fail unexpectedly. Thus, in order to make a fair comparison, the variances should be noted.

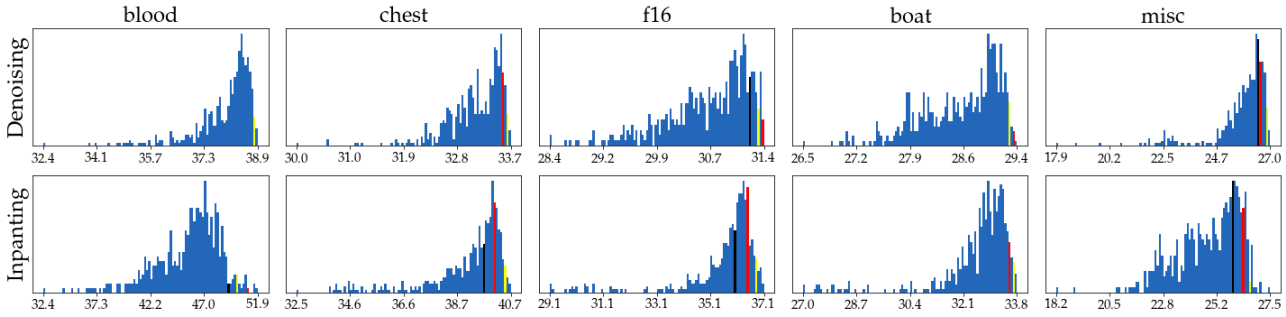


Figure 6: Histograms of the final PSNR scores of models for all images in *the NAS Dataset for DIP* for image denoising and image inpainting . Red and black colored bars represent the final PSNR scores of the best models included in top-N models found by PSD DB Strip MSE metric. The bin depicted with red denotes the best model among N=15 models, the black bin denotes the best one among N=5 models. The bin depicted with yellow denotes the average PSNR scores (which was repeated 10 times) of the models each of which was the best performing model among randomly sampled 15 models.

References

[1] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1, 4

[2] Yun-Chun Chen, Chen Gao, Esther Robb, and Jia-Bin Huang. Nas-dip: Learning deep image prior with neural architecture search. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 442–459, Cham, 2020. Springer International Publishing. 1, 4