# Natural Image Noise Dataset

Benoit Brummer and Christophe De Vleeschouwer

Université Catholique de Louvain, Belgium

## Introduction

Photographic image noise occurs when a camera is not given enough time to capture the information present in a scene with the given settings. The sensor sensitivity (ISO setting) is increased and the result is akin to the sensor making random guesses; a noisy image has been captured.

Photographic development involves the use of a denoising algorithm to detect the noise and reconstruct the latent clean image. Convolutional neural networks have been the focus of research aiming to solve the denoising problem, but they require large amount of paired data to learn a mapping. In the absence of such data, artificial noise is added to a clean image and the result is unsatisfactory in real life applications.

We introduce the Natural Image Noise Dataset; a photographic image dataset of clean-noisy scenes taken with different ISO settings that is large, diverse, and of high enough quality to train neural networks for image denoising tasks.

We obtain state-of-the-art results with a single trained generative model, using various network architectures (including RED-Net and U-Net). The models appears to generalize well to other types of cameras.

## Methods

### Capture

We captured several images per static scene; at least one ground-truth taken with the camera's lowest ISO setting and several images taken with increasing ISO settings and consequent decreasing shutter speed in order to match the original exposure value. Scenes were captured using a camera affixed to a tripod and controlled with a wireless remote control to avoid shifting the setup position. We ensured that the ground was stable, wind would not cause any change in the scene, lighting did not vary between shots, and no area was overexposed in the ground-truth images. Overexposure occurs when the sensor is saturated. On a ground-truth image this would potentially benefit the high ISO images because less light is captured with a faster shutter speed; therefore, the higher ISO images may not be overexposed and thus may contain detail that is not present in the overexposed ground-truth. The aperture remained the same on all shots and the focus was set manually so that it would not automatically adjusted for each frame.

Several images were taken with different intermediate ISO values such that the ISO settings varied across each scene. We often also took images that we categorized as "High ISO," which consisted of the highest ISO value and increased shutter speed. "High ISO" images result in dark frames which are then correctly exposed using software. We often tried to match shutter speeds that would be useful to denoise, such as 1/60 s for handheld photography, 1/15 s for devices equipped with optical image stabilization, and 1/1000 s or faster for high-speed photography. We ensured that every ISO value was well represented in order to train models effective at blind denoising. On average, six images were produced per scene and some scenes featured multiple ground truth images which could be used in training to help prevent over-fitting. The dataset currently consists of 616 images from 101 different scenes. A large subset of the ground-truth images is shown on Figure 1.

### Software processing

We processed the dataset images using darktable (an open source development software) for raw-to-sRGB development. Our development steps are similar to those we would apply to a standard picture but no sharpening was applied, as this greatly amplifies noise and is typically applied last in the pixel pipeline (we can expect users to apply sharpening to the generated clean image without any perceptible loss). We used darktable's automatic exposure mode to match a fixed percentile target on the histogram and calculate the required exposure compensation for all images in a set. Likewise, we ensured that the white balance was identical and all development steps were copied over to the entire scene. (The exposure percentile and white balance settings are fixed within a scene but vary across different scenes.) The raw overexposed indicator was used to verify that no overexposed areas were present in the base ISO image and if any were detected then it was cropped out or the scene was discarded. The images were visually inspected to detect slight variations in light, the introduction of foreign objects such as insects, or any movement, which also resulted in cropping or discarding images. The ground-truth images must be at least as sharp as their noisy counterparts; this is sometimes not the case due to slight movements on longer exposures. The remaining images were saved in either high-quality (98 to 100) 8-bit JPEG or lossless 16-bit PNG.

The last step of development is to use Hugin's align_image_stack tool to ensure that all images in a set are perfectly pixel-aligned. The tool will usually return the same image size as the input, in which case the whole image set can safely be used. When a difference is detected then the tool will automatically align the set and we visually analyzed whether the result was acceptable or the movement caused a change in perspective, in which case the outlier images were discarded. Some noisy images cannot easily be matched to the scene; possible solutions are to denoise these images in order to check the alignment or to take a cleaner image afterward and assume that the middle images are consistent with the previous and next ones.

### Publication

The Natural Image Noise Dataset is published on Wikimedia Commons, an online repository of free-use images and other digital media.
One key advantage of using Wikimedia Commons is its collaborative aspect. Anyone is allowed to add images to the dataset, modify existing images (for example to fix a spatial misalignment), and discuss the content (through the discussion page provided for each file, category, and the dataset itself). The collaborative aspect also includes a "Quality images candidates" page where users assess the technical quality of a submitted image and may promote it to a "Quality image" standing. Many of the ground-truth images have gone through this process and were promoted through human assessment.

Thumbnails are generated and images may be visualized before being downloaded in full resolution, and the download may include a select subset instead of the whole dataset. A customizable download script is provided on the dataset's page for convenient retrieval. Even though files can be overwritten, every file uploaded on Wikimedia Commons is kept forever therefore specific snapshots of the dataset can be made and the download script can fetch the dataset as it was on any specified date.

## Availability

The NIND is available for download and third-party contributions on:
**https://commons.wikimedia.org/wiki/Natural_Image_Noise_Dataset**

Figure 1: Sample images from the NIND (center) with visual noise comparison (edges)

## Results

We trained various network architectures (such as DnCNN, RED-Net, and U-Net) using the NIND and found that most networks obtain state-of-the-art performance, alas some networks provide significantly faster training and inference. We ended up using U-Net for most of our experiments as it provides a useful inference speed of at most 0.75 s per megapixel on a GeForce GTX 1070. The model's performance is shown on Figure 2 (SSIM), Figure 4 (visually), and Figure 3 (both). We also acquired images with a different camera to test the generalization performance and found that the model trained on one camera could generalize without percpetible performance hit compared with the model trained on both cameras. (Figure 5) We also tested whether adding different data could deteriorate performance by adding 320 images from the SIDD and we found that the model could handle them as well (although a model trained only on the SIDD did not perform as well on our DSLR-like images).

Figure 2:
Test set denoising performance of
U-Net model trained on NIND,
U-Net trained on artificial noise,
BM3D
On five test images

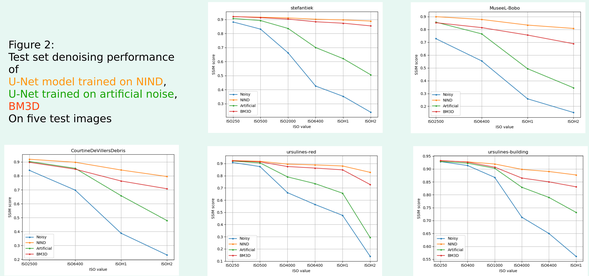

Figure 3: denoising stefantiek. 1: ISO200 ground-truth (1/20 s, SSIM: 1.00), 2: high ISO (1/1900 s, SSIM: 0.239), 3: 2 denoised using U-Net model trained with NIND (SSIM: 0.890), 4: 2 denoised using U-Net model trained with SIDD-320 (SSIM: 0.870), 5: 2 denoised using BM3D (SSIM: 0.854)

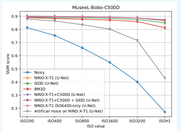


Figure 5: Generalization performance comparison: Image from a Canon 500D camera denoised with models trained on the NIND with only Fujifilm X-T1 pictures vs the entire dataset.

Figure 4: Visual comparison of a high-speed low-aperture ISO6400 image (top) denoised with a U-Net model trained with the NIND (middle) and with BM3D (bottom)

### Code

Our code is available on **https://github.com/trougnouf/mthesis-denoise**
This includes the download script, a cropping script, a PyTorch dataset handler, a training script with various architectures (such as U-Net, PatchGAN, and our own experimental "Hulb" architectures) and support for both single generative networks (SSIM and/or L1-based) and (conditional) generative adversarial networks (with currently up to two different discriminators), an image denoising script and a directory-based test suite.

## Bibliography

NIND
U-Net
Pix2pix
Darktable
SIDD