DEEP CLASS-AWARE IMAGE DENOISING

Tal Remez[†] *Or Litany*[†] *Raja Giryes*[†] *Alex M. Bronstein*[⋆†]

[†] School of Electrical Engineering, Tel-Aviv University, Israel
[⋆] Computer Science Department, Technion - IIT, Israel

ABSTRACT

The increasing demand for high image quality in mobile devices brings forth the need for better computational enhancement techniques, and image denoising in particular. To this end, we propose a new fully convolutional deep neural network architecture which is simple yet powerful and achieves state-of-the-art performance for additive Gaussian noise removal. Furthermore, we claim that the personal photo-collections can usually be categorized into a small set of semantic classes. However simple, this observation has not been exploited in image denoising until now. We show that a significant boost in performance of up to 0.4dB PSNR can be achieved by making our network class-aware, namely, by fine-tuning it for images belonging to a specific semantic class. Relying on the hugely successful existing image classifiers, this research advocates for using a class-aware approach in all image enhancement tasks.

Index Terms— Image denoising, machine learning, computer vision, image enhancement, image processing.

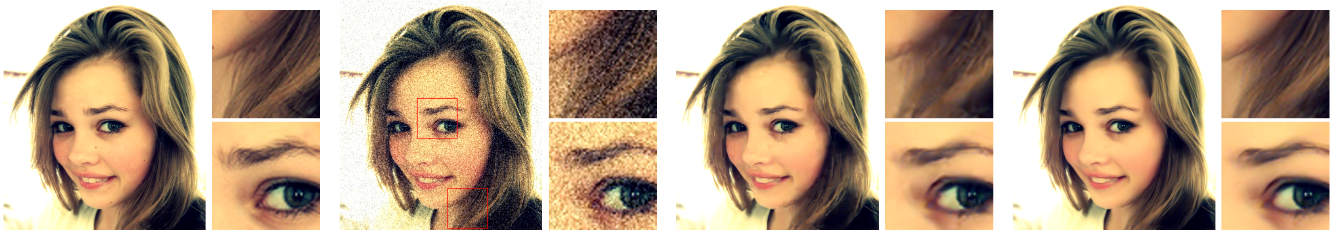
1. INTRODUCTION

Many image acquisition artifacts such as low-light noise and camera shake can be compensated by image enhancemnet techniques. Denoising in the presence of additive white Gaussian noise is one of the key problems studied in this context. It has been shown in [2, 3, 4] that having a good Gaussian denoising algorithm allows to efficiently solve many other image processing problems such as deblurring, inpainting, compression postprocessing, low-light Poisson denoising and more, without compromising the reconstruction quality or the need to design a new strategy adapted to a new setting. Numerous methods have been proposed for removing Gaussian noise from images, including k-SVD [5], non-local means [6], BM3D [7] , field of experts (FoE)[8] and many others. These techniques were designed based on some properties of natural images such as the recurrence of patches at different locations and scales, or their sparse representation in some (possibly trained) dictionary. In the past few years, the state-of-the-art in image denoising has been achieved by techniques based on artificial neural networks. The first such method was the multilayer perceptron (MLP) proposed in [9]. Which is based on a fully connected architecture and therefore requires a very large amount of training examples, memory and has high arithmetic complexity at inference compared to e.g. the recent work [10], which proposes a neural network based on a deep Gaussian Conditional Random Field (DGCRF) model, or the model-based Trainable Nonlinear Reaction Diffusion (TNRD) network introduced in [1].

Class Aware Denoising. Patch-based image denoising theory suggests that existing methods have practically converged to the theoretical bound of the achievable performance [11]. As it turns out, two possibilities to break this barrier still exist. The first is to use larger patches. This has been proved useful in [9] where the use of 39×39 patches allowed to outperform BM3D. The second is to use a better image prior, such as narrowing down the space of images to a more specific class. These two possibilities are not mutually exclusive, and indeed we exploit both. Several studies have shown that it is beneficial to design a strategy for a specific class. In [12], the authors set a bound on super-resolution performance and showed it can be broken when a face-prior is used. In [13], a compression algorithm for face images was proposed. Face hallucination, super-resolution and sketch-photo synthesis methods have been developed by [14]. In [15], the authors showed that given a collection of photos of the same person it is possible to obtain a more faithful reconstruction of the face from a blurry image. In [16, 17] class labeling at a pixel-level was used for the colorization of gray-scale images.

Contribution. We propose a novel convolutional neural network (CNN)-based architecture that obtains performance higher than or comparable to the state-of-the-art for Gaussian image denoising. We demonstrate that an additional boost in performance is achieved when the algorithm is aware of the semantic class of images being processed, or *class-aware*. Different from previous methods, our model is made class-aware via training and not by design. While in this paper we focus on Gaussian denoising, our methodology can be easily extended to much broader class-aware image enhancement, rendering it applicable to many computational photography and low-level computer vision tasks.

978-1-5090-2175-8/17/$31.00 ©2017 IEEE



Ground truth image
Noisy image
Denoised by TNRD [1]
Denoised by our method

Fig. 1. **Perceptual comparison of class-aware and standard denoising.** Our proposed face-specific denoiser produces a visually pleasant result and avoids artifacts commonly introduced by general-purpose denoisers. The reader is encouraged to zoom in for a better view of the artifacts.

2. DENOISE NET

Our network performs additive Gaussian image denoising in a fully convolutional manner. It receives a noisy grayscale image as the input and produces an estimate of the original clean image. The network architecture is shown in Figure 2. The layers at the top row of the diagram calculate features using convolutions of size 3×3 , stride 1, and *ReLU* non-linearities. While the layers at the bottom of the diagram can be viewed as negative noise components as their sum cancels out the noise, and are calculated using a single channel convolution of size 3×3 with stride 1. In all experiments we used networks with 20 layers implemented in TensorFlow [18] and trained it for $160K$ mini-batches on a Titan-X GPU with a set of 8000 images from the PASCAL VOC dataset [19]. We used mini-batches of 64 patches of size 128×128. Images were converted to YCbCr and the Y channel was used as the input grayscale image after being scaled and shifted to the range of $[-0.5, 0.5]$. During training, image patches were randomly cropped and flipped about the vertical axis. To avoid convolution artifacts at the borders of the patches caused by the receptive field of pixels in the deepest layer, we used an ℓ_2 loss on the central part cropping the outer 21 pixels during training time and padded the image symmetrically during test time by 21. Training was done using the ADAM optimizer [20] with a learning rate of $\alpha = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. Code and pretrained models will be made available[1].

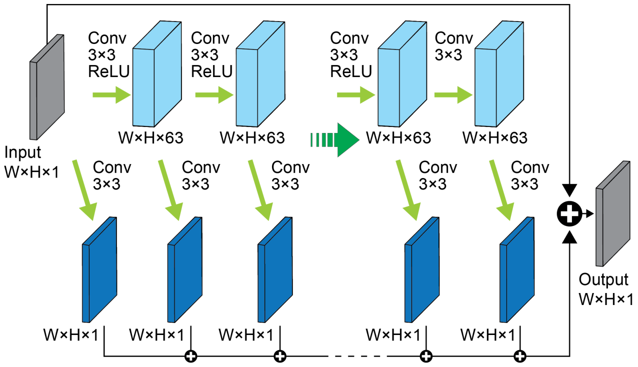


Fig. 2. **DenoiseNet fully convolutional architecture.** All convolutions are of size 3×3 and stride 1. Convolution resulting feature sizes are listed as $Width \times Height \times \#Channels$. The bottom row of outputs can be viewed as a negative noise components as their sum cancels out the noise.

3. CLASSIFICATION IN THE PRESENCE OF NOISE

The tacit assumption of our class-aware approach is the ability to determine the class of the noisy input image. While the goal of this research is not to improve image classification, we argue that the performance of modern CNN based classification algorithm such as *Inception* [21, 22] or *ResNet* [23] is relatively resilient to a moderate amount of noise. In addition, since we are interested in canonical semantic classes such as *faces* and *pets* which are far coarser than the 1000 ImageNet classes [24], the task becomes even easier. Furthermore, the aforementioned networks can be fine-tuned using noisy examples to increase their resilience to noise. To illustrate the noise resilience property we ran the pre-trained *Inception-v3* [22] network on a few tens of images from the *pets* class. We then gradually added noise to these images and counted the number of images on which the classifier changed its label to non-pet. Observe that in Figure 3 the network classification remains stable even in the presence of large amounts of noise.

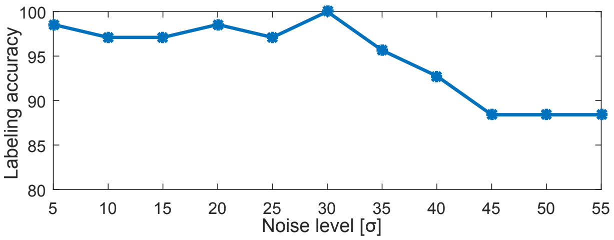


Fig. 3. **Noise resilience of image classification.** Correct classification rate of *inception-v3* in the presence of noise.

[1] https://github.com/TalRemez/deep_class_aware_denoising

4. EXPERIMENTS

In all experiments our network was compared to (a) BM3D ;(b) multilayer perceptrons (MLP) using sigma specific, pre-trained models trained on ImageNet and made available online; and (c) TNRD using sigma specific, pre-trained models published by the authors; on the following test sets: (i) images from PASCAL VOC [25]; and (ii) the commonly used 68 test images chosen by [8] from the Berkeley segmentation dataset [26].

4.1. Class-agnostic denoising

In all experiments in this section our network was trained on $8K$ images from the PASCAL VOC [25] dataset.

PASCAL VOC. In this experiment we tested the denoising algorithms on $1K$ test images from [25]. Table 1 summarizes performance in terms of average PSNR for white Gaussian noise with σ ranging from 10 to 75. It is evident that our method outperforms all other methods for all noise levels by a significant margin.

σ	10	15	25	35	50	65	75
BM3D	34.26	32.10	29.62	28.14	26.61	25.64	25.12
MLP [9]	34.29	–	29.95	28.49	26.98	26.07	25.54
TNRD [1]	–	32.35	29.90	–	26.91	–	–
DenoiseNet	**34.87**	**32.79**	**30.36**	**28.88**	**27.32**	**26.30**	**25.74**

Table 1. **Performance on PASCAL VOC.** Average PSNR values on a 1000 image test set. Our method outperforms all other methods for all noise levels.

To examine the statistical significance of the improvement our method achieves for $\sigma = 25$, in Figure 4 we compare the gain in performance with respect to BM3D achieved by our method, MLP and TNRD. The plot visualizes the large and consistent improvement in PSNR achieved by our method and that it outperforms all others on 92.4% of the images, whereas MLP, BM3D, and TNRD win on $6.6\%, 1\%$ and 0% respectively.

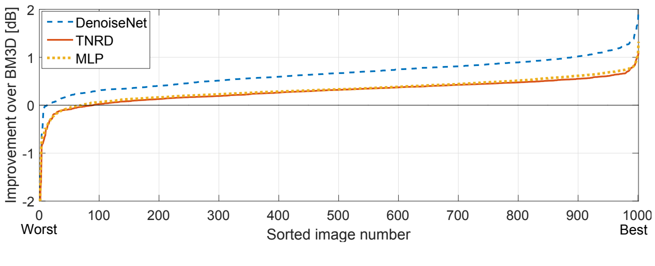


Fig. 4. **Performance profile relative to BM3D.** Image indices are sorted in ascending order of performance gain relative to BM3D. The improvement of our method is demonstrated by (i) decrease of the zero-crossing point, and (ii) consistently higher values of gain. The comparison was made on images from PASCAL VOC for $\sigma = 25$.

Berkeley segmentation dataset. In this experiment we tested the performance of our method, trained on PASCAL VOC, on the widely used test-set of 68 images selected by [8] from Berkeley segmentation dataset [26]. Even though these test images belong to a different dataset, Table 2 shows that our method outperforms previous methods for all σ values.

σ	10	15	25	35	50	65	75
BM3D	33.31	31.10	28.57	27.08	25.62	24.68	24.20
MLP [9]	33.50	–	28.97	27.48	26.02	25.10	24.58
TNRD [1]	–	31.41	28.91	–	25.95	–	–
DenoiseNet	**33.58**	**31.44**	**29.04**	**27.56**	**26.06**	**25.12**	**24.61**

Table 2. **Performance on images from Berkeley segmentation dataset.** Average PSNR values on a test set of 68 images selected by [8]. Our method outperforms all others for all noise levels.

4.2. Class-aware denoising

This experiment evaluates the boost in performance gained by fine-tunning a denoiser on a set of images belonging to a particular class. In order to do so we collected images from ImageNet [24] of the following six classes: *face*, *pet*, *flower*, *beach*, *living room*, and *street*. The $1,500$ images per class were split into train (60%), validation (20%) and test (20%) sets. We then trained a separate class-aware denoiser for each of the classes for a noise level of $\sigma = 25$. This was done by fine-tuning all the parameters of our class-agnostic model, that had been trained on PASCAL VOC (and used in section 4.1), using the images from ImageNet. The performance of the class-aware denoisers was compared to their class-agnostic counterpart and to other denoising methods (using their online published coed and models). Aver-

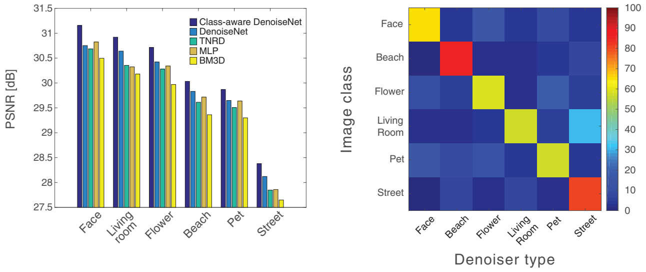


Fig. 5. **(Left) Class-aware denoising on ImageNet.** Average PSNR values on images belonging to six different semantic classes. It is evident that the class-specific fine-tuned models outperform all other methods. In addition, being class-aware enables to gain as much as 0.4 dB PSNR compared to our class-agnostic network. **(Right) Cross-class denoising.** Each row represents a specific semantic class of images while class-aware denoisers are represented as columns. The (i, j)-th element in the confusion matrix shows the probability of the j-th class-aware denoiser to outperform all other denoisers on the i-th class of images.

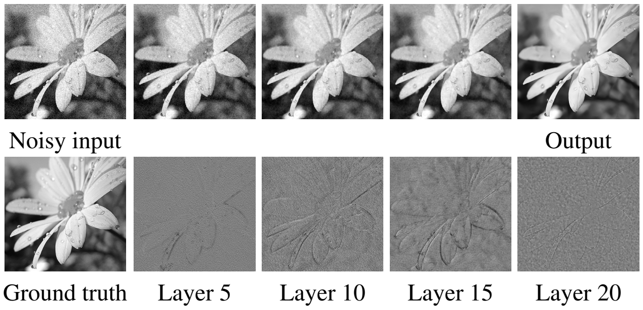


Fig. 6. **Gradual denoising process by *flower*-specific DenoiseNet.** The top row presents the noisy image (left) and the intermediate result obtained by removing the noise estimated up to the respective layer depth. The second row presents the ground truth image (left) and the noise estimates produced by individual layers; the noise images have been scaled for display purposes. We encourage the reader to zoom-in onto the images to best view the fine details and noise.

age PSNR values summarized in Figure 5 indicate that our class-aware models outperforms our class-agnostic network, as well as BM3D, MLP and TNRD; demonstrating a boost in performance by as much as $0.4dB$. An extensive qualitative comparison is available in the supplementary material.

Cross-class denoising To further demonstrate the effect of refining a denoiser to a particular class, we tested each class-specific denoiser on images belonging to other classes. To quantify the effect of mismatching we evaluated the percentage of wins of every fine-tuned denoiser on each type of image class. A win means that a particular denoiser produced the highest PSNR among all the others. Figure 5(Right) shows a confusion matrix for all combinations of class-specific denoisers and image classes. Qualitatively it is clear our class-specific denoisers learn to better handle textures, edge types, and structures that are common in each of the classes, and that using the wrong denoiser creates artifacts that resemble the statistics of textures and edges from the densoier class. We refer the reader to the supplementary material for examples. We conclude that applying a denoiser of the same class as the image results in the best performance.

4.3. Network noise estimation

In order to gain some insights about our network's noise estimation process, in Figure 7 we show the error after $5, 10$, and 20 layers (middle row). Surprisingly, even thought it has not been explicitly enforced at training, the error monotonically decreases with the layer depth (plot at the bottom left). This non-trivial behavior is consistently produced by the network on most test images. By visualizing which layer was the most dominant in the denoising process of each pixel we observe that the first few layers govern the majority of smooth image areas, while the deeper layers contribute the most to edges (bottom right). In addition, as demonstrated in Figure 6, each layer of the network contributes differently to the noise removal process. The shallower layers seem to handle local noise statistics while the deeper layers recover edges and enhance textures that might have been degraded by the first layers. A possible explanation for this may reside in their receptive field sizes. Deeper layers correspond to larger receptive fields, therefore can better recover large patterns such as edges, contours, and textures, which might be indistinguishable from noise when viewed by smaller receptive fields of shallower layers.

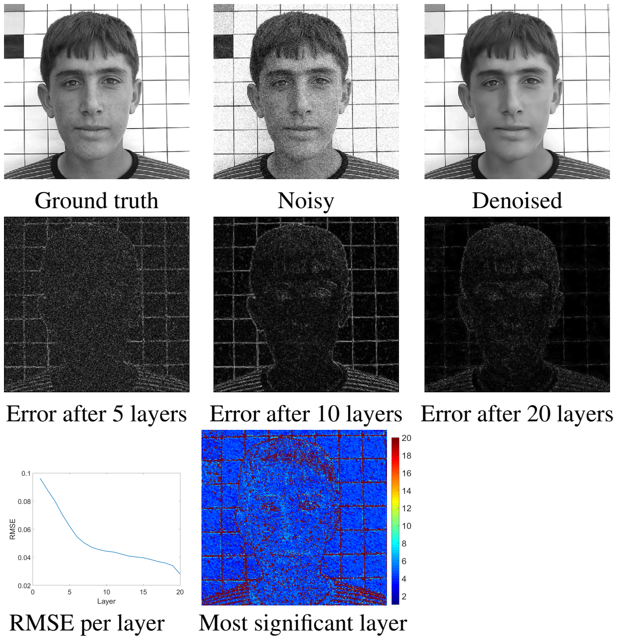


Fig. 7. **Gradual denoising process.** Top row: ground truth, noisy ($\sigma = 25$) and denoised images. Middle: difference images evaluated after accumulating noise estimations from the first $5, 10$, and 20 layers. Bottom left: RMSE at each layer. Bottom right: To visualize which of the layers was the most dominant in the denoising process, we assigned a different color to each layer according to it's depth, and colored each pixel according to the layer in which its value changed the most. Images are best viewed electronically, one should zoom in for a better view. More examples are available in the supplementary material.

5. DISCUSSION

We introduced a new fully convolutional neural network for image denoising with state-of-the-art performance. We further showed that fine-tuning the network per class is preferable over a universal filter, achieving an additional boost of up to 0.4dB PSNR. That said, the decision to split according to a semantic class was made due to the immediate availability of off-the-shelf classifiers and their resilience to noise. Yet, this splitting scheme may very well be sub-optimal and one could instead incorporate it into the network architecture and refine via end-to-end training. We defer this to future research.

6. REFERENCES

[1] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence (CVPR)*, 2016.

[2] Mauricio Delbracio and Guilermo Sapiro, "Burst deblurring: Removing camera shake through fourier burst accumulation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[3] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *arXiv:1611.02862*, 2016.

[4] S.V. Venkatakrishnan, C.A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *GlobalSIP*, 2013.

[5] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[6] J.M. Morel A. Buades, B. Coll, "A non-local algorithm for image denoising," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[7] V. Katkovnik K. Dabov, A. Foi and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[8] Stefan Roth and Michael J Black, "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009.

[9] C. J. Schuler H. C. Burger and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2392–2399.

[10] O. Tuzel R. Vemulapalli and M. Liu, "Deep gaussian conditional random field network: A model-based deep network for discriminative denoising," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[11] F. Durand A. Levin, B. Nadler and W.T. Freeman, "Patch complexity, finite pixel correlations and optimal denoising," in *ECCV*, 2012.

[12] S.Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[13] Ori Bryt and Michael Elad, "Compression of facial images using the K-SVD algorithm," *Journal of Visual Communication and Image Representation*, vol. 19, no. 4, pp. 270 – 282, 2008.

[14] Nannan Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li, "A comprehensive survey to face hallucination," *International Journal of Computer Vision*, vol. 106, no. 1, pp. 9–30, 2014.

[15] Neel Joshi, Wojciech Matusik, Edward H. Adelson, and David J. Kriegman, "Personal photo enhancement using example images," *ACM Trans. Graph.*, vol. 29, no. 2, pp. 12:1–12:15, Apr. 2010.

[16] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," in *SIGGRAPH*, 2016.

[17] Richard Zhang, Phillip Isola, and Alexei A Efros, "Colorful image colorization," *ECCV*, 2016.

[18] P. Barham E. Brevdo Z. Chen C. Citro G. S. Corrado A. Davis J. Dean M. Devin M. Abadi, A. Agarwal et al., "Tensorflow: Large-scale machine learning on heterogeneous systems, 2015," *tensorflow.org*.

[19] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[20] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.

[21] Y. Jia P. Sermanet S. Reed D. Anguelov D. Erhan V. Vanhoucke C. Szegedy, W. Liu and A. Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015.

[22] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision, journal = arXiv, abs/1512.00567, year = 2015, url = http://arxiv.org/abs/1512.00567,," .

[23] S. Ren J. Sun K. He, X. Zhang, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[24] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *Int. Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[25] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results," http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html.

[26] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, July 2001, vol. 2, pp. 416–423.