Ground-Truth Data Set and Baseline Evaluations for Base-Detail Separation Algorithms at the Part Level

Xuan Dong, Boyan I. Bonev, Weixin Li, Weichao Qiu, Xianjie Chen, and Alan L. Yuille, *Fellow, IEEE*

Abstract—**Base-detail separation is a fundamental image processing problem, which models the image by a smooth base layer for the coarse structure and a detail layer for the texturelike structures. Base-detail separation is hierarchical and can be performed from the fine level to the coarse level. The separation at coarse level, in particular at the part level, is important for many applications, but currently lacks ground-truth data sets that are needed for comparing algorithms quantitatively. Thus, we propose a procedure to construct such data sets and provide two examples: Pascal Part UCLA and Fashionista, containing 1000 and 250 images, respectively. Our assumption is that the base is piecewise smooth, and we label the appearance of each piece by a polynomial model. The pieces are objects and parts of objects obtained from human annotations. Finally, we propose a way to evaluate different separation methods with our data sets and compared the performances of seven state-of-the-art algorithms.**

Index Terms—**Base-detail separation, part level.**

I. Introduction

BASE-DETAIL separation is a fundamental problem in image processing, and is useful for a number of applications, such as contrast enhancement [1], exposure correction [2], and so on. It defines a simplified coarse representation of an image with its basic structures (base layer), and a detailed representation, which may contain texture, fine details, or just noise (detail layer), as shown in Fig. 1.

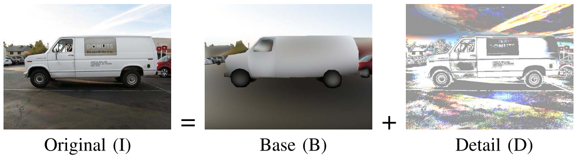

Fig. 1. Base-detail separation is a fundamental image processing problem. It relates to a wide range of tasks, such as contrast enhancement, exposure correction, and so on. (Here, D is shown with increased contrast for clarity.)

This definition leaves open what is detail and what is base. We argue that base-detail separation should be formulated as a hierarchical structure. For instance, in an image of a crowd of people, we argue that their heads and faces form a texture, or detail, over a common base surface, which could be their average color. At a less coarse level, we could say that each individual head is composed of two base regions: the hair and the face whereby the details are the hair texture, the eyes, nose, and mouth. We could go into still more detail and argue that the mouth of a person could also be separated into a smooth surface (or base), and the details of the lips, if there is enough resolution. Roughly speaking, from fine to coarse, the hierarchical base-detail separation can be classified as the pixel level, the subpart level, the part level, and the object level. Fig. 2 shows an example of hierarchical base-detail separation of an image.

Benchmark data sets are becoming important in computer vision and image processing. For most computer vision problems, such as optical flow [3], stereo [4], object recognition [5], and edge detection [6], there exist data sets used as benchmarks for evaluation and comparison. These data sets have driven innovation and rigor to those problems. However, there is a lack of a common data set for image base-detail separation at the coarser levels of the hierarchy, in particular, at the part level. This makes it difficult to draw conclusions and to compare quantitatively.

Base-detail separation at the part level is important for many applications and failing to do it correctly will introduce artifacts into the final enhancement results. Fig. 4 shows the types of errors, such as halo artifacts, in exposure-correction enhancement [2] that result from incorrect separation. Some examples of the incorrect separation are shown in Fig. 3.

Thus, we generate a ground-truth base-detail data set at the part level in this letter. At the part level, a fundamental assumption of this letter is that the base layers are piecewise smooth. It is piecewise because of the different objects and parts present in the image. For each part at the part level, the separation results should not be affected by their neighboring parts, hence the methods should successfully preserve the sharp boundaries. Otherwise, halo artifacts will be introduced into both of the base and the detail, as shown in Fig. 3.

To get the ground-truth data, we manually segment each image into parts, and for each part, we label the base and detail layers. Segmenting images into parts is challenging because the shapes of parts and their appearances vary a lot. We rely on manual annotations of the segments at the pixel level. Within each part, labeling the base-detail separation is also challenging, because it requires pixelwise annotation for the intensity of the base layer in the RGB color space (or, more generally, in any color space). We use a polynomial model of different orders to separate the image signal into several possible base layers and let humans select which order of the polynomial separation is the correct base layer separation. The residual of the selected base layer is the detail. It is possible that none of the polynomial model's results is correct for the base layer. So, in the annotation, we exclude the images if even one region of the image cannot be described by the polynomial model.

The main contributions of this letter are: 1) two natural image data sets providing the part-level base and detail ground truth (Pascal

Manuscript received April 25, 2016; revised August 2, 2016 and October 2, 2016; accepted October 8, 2016. Date of publication October 19, 2016; date of current version March 5, 2018. This work was supported by NSF under Grant NSF CCF-1317376. This paper was recommended by Associate Editor A. Loui. *(Xuan Dong and Boyan I. Bonev contributed equally to this work.) (Corresponding author: Weixin Li.)*

X. Dong, B. I. Bonev, W. Li, W. Qiu, and X. Chen, are with the Department of Statistics, University of California, Los Angeles, Los Angeles, CA 90095 USA (e-mail: dongx@ucla.edu; boyan@ua.es; lwx@cs.ucla.edu; qiuwch@gmail.com; cxj@ucla.edu).

A. L. Yuille is with the Department of Statistics, University of California, Los Angeles, Los Angeles, CA 90095 USA, and also with the Department of Cognitive Science and Computer Science, John Hopkins University, Baltimore, MD 21218 USA (e-mail: alan.l.yuille@gmail.com).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2016.2618933

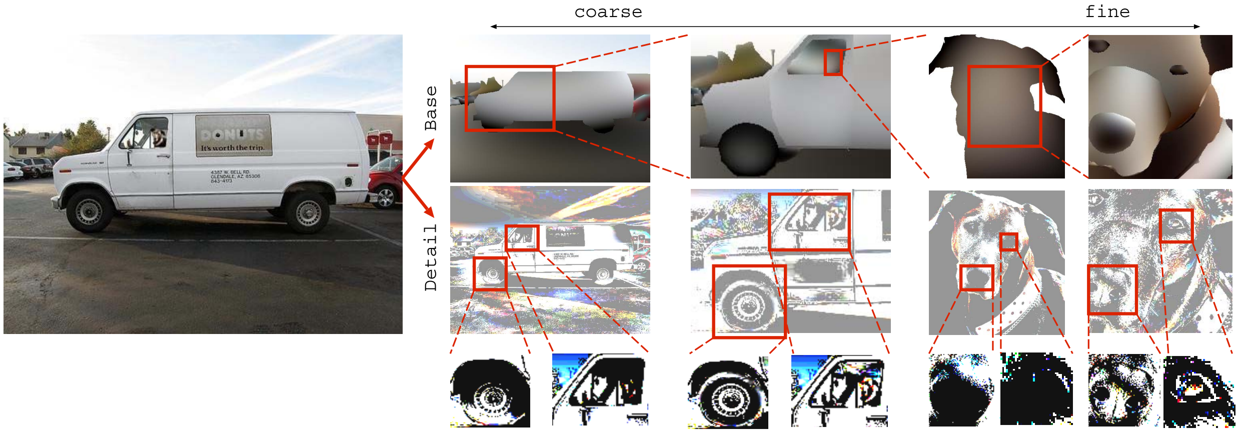

Fig. 2. Example of hierarchical base-detail separation. The original image (left) is decomposed into base (first row) and detail (second row). The columns are sorted from coarser to finer taxonomy in the hierarchy. Depending on the hierarchy level, details contain different information, e.g., the nose of the dog can be considered as detail, or it can be considered as a part, and the nostrils are the detail (see third row). Note that detail is represented with a very high contrast here, for better visualization.

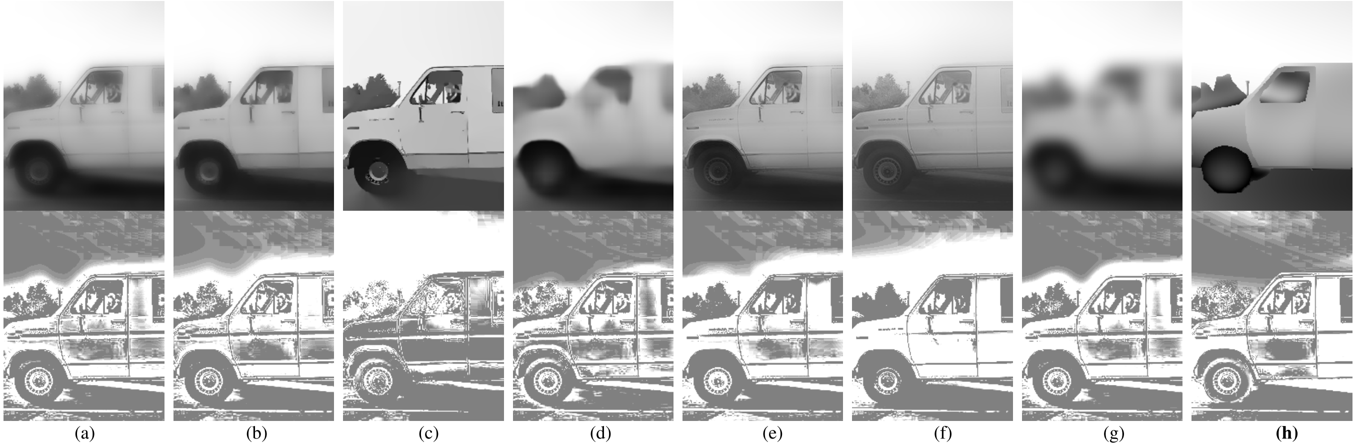

Fig. 3. Halo effects produced by different filters [(a)–(g)] compared with proposed ground truth [(h)]. Top row: base. Bottom row: detail, with increased contrast for visualization. (a) AM [8], (b) DT [7], (c) L0 [1], (d) RG [9], (e) BL [10], (f) GD [11], and (g) GS. The different halo effects are easily observed between the sky and the vehicle, and most of them produce blurry boundaries in the base layer and a padding artifact in the detail layer. Another effect is that some methods represent the cloud as a broken texture (detail layer), and not as a bloblike structure [(h) ground truth detail].

Part UCLA and Fashionista) and 2) the evaluation of several state-of-the-art base-detail separation algorithms based on the provided ground truth. The supplementary materials include more detailed information of the related work and this letter, and more experimental results.

II. Related Work

A. Image Base-Detail Separation Algorithms and Applications

We argue that the separation of image base-detail layers is hierarchical. And different image base-detail separation algorithms focus on separation at different levels of the hierarchical structure, aiming at different applications.

Base-detail separation at fine level usually aims at the separation of signal and noise. The detail layer is occupied by noise, whereas the base layer contains the image signal. Different algorithms have been proposed for applications, such as denoising [12], joint upsampling [13], compression artifact removal [1], and so on.

Base-detail separation at coarse level aims to separate the image's basic coarse structures, such as the luminance and color of parts, and texturelike detail, such as high-frequency and midhigh-frequency information. Different algorithms are proposed for different applications, such as exposure correction [2], HDR/tone mapping [14], style transfer [15], tone management [16], contrast enhancement [7], image abstraction [1], and so on.

B. Related Data Sets and Quantitative Evaluation

There are some quantitative evaluation methods for base-detail separation at the fine level, suitable for image denoising [12], upsampling [13], and compression artifact removal [1]. But few works have been proposed for quantitative evaluation at the part level. The performance of base-detail at the fine level has little relationship with the performance at the part level because ignoring the piecewise smoothness assumption does not have big effects on the results at fine level but can have a very big effect at the part level. For example, at the part level, artifacts like halos are frequently introduced because the separation of a part is contaminated by neighboring parts. Thus, ground-truth base-detail separation at the part level is lacking and desired.

III. Ground-Truth Base-Detail Separation Data Set

The goal of this letter is to construct a ground-truth base-detail separation data set at the part level. We assume that the separated base layer is piecewise smooth (validated by our human annotators). It means that the separation of pixels of one part is determined only by the part itself, and the neighboring parts do not affect it. This avoids the halo effects between parts (see Figs. 3 and 4). Thus, to get the ground-truth base-detail separation data set, we propose to use images that are manually segmented at the part level, and within each part, we annotate the base layer in the RGB color space by using a polynomial model to separate each part into several possible base layers and letting humans select the correct one.

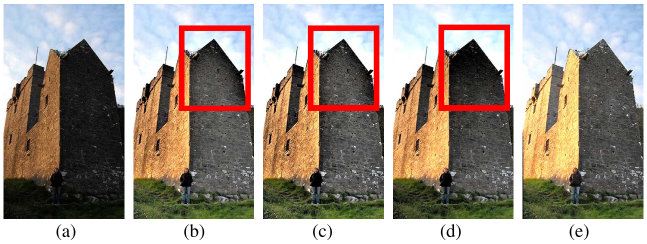

Fig. 4. An example showing halo effects produced by different filters for exposure correction. (a) The input image. (b) Result of the AM filter. (c) Result of the DT filter. (d) Result of the GD filter. (e) Result using the proposed ground-truth base-detail separation result. The halo effects are easily observed between the sky and the building (as marked in the red box region), which are caused by incorrect base-detail separation at the part level. For more examples, see the supplementary material.

We rely on pixelwise human annotations for the part-level segmentation because the shapes of different parts vary a lot, and pixelwise annotations could get accurate segmentation results. There exist many automatic segmentation algorithms [17], but they are not accurate enough for our work. For base-detail separation within each part, we use both polynomial and human annotation, because it is difficult for humans to directly annotate the base and detail layers of every pixel in the RGB color space. Thus, we propose to first separate the image signals of each part into several possible results. Then, from the set of possible separation results, we let humans select which one gives the best base-detail separation of each part. We use a polynomial model with different orders, which is one of the most basic signal processing methods, to produce several possible results for each part. This strategy reduces the labeling work significantly and makes it much easier to generate the ground-truth data sets for base-detail separation at the part level.

A. Annotation for Segmentation

The annotation for segmentation at the part level consists of drawing the exact boundaries of each part in the image. Fortunately, there exist some well-known data sets with this type of annotations, such as the Fashionista [18] and the Pascal Part UCLA data set [19]. They provide human-made annotations of parts at the pixel level. We reuse their part-level annotations. In Fashionista, the labeled regions are parts of human beings, such as hair, glasses, and so on, and the segmented parts have appearance consistency. In the Pascal Part UCLA data set, the labeled images have many classes, such as dog, person, and so on. And regions are labeled according to semantic meaning and do not necessarily enforce the appearance consistency.

The definition of a part is different in different applications and data sets, because the base-detail separation is a hierarchical structure. This explains why the Fashionista and the Pascal Part UCLA data sets used different strategies for parts annotation. In our opinion, this diversity of the labeling of parts is a good property because we can evaluate different base-detail separation algorithms at different levels of the hierarchical structure so that we can have a better understanding of the performance of different algorithms.

B. Annotation for Base-Detail Separation

To annotate the base layer within each part, the first step is to separate the image signals of each part into several possible base layers with the polynomial model at different orders. Within each part P_i, we fit polynomial models on each color channel separately. The number of parameters $\vec{\omega}$ depends on the order k of the polynomial. The polynomial approximations are

$$\begin{aligned} b_k(\vec{x}, \vec{\omega}) &= \vec{x}^T \vec{\omega} \\ k = 0 : \vec{x} &= 1, \vec{\omega} = \omega_0 \\ k = 1 : \vec{x} &= [1, x_1, x_2], \quad \vec{\omega} = [\omega_0, \omega_1, \omega_2] \\ k = 2 : \vec{x} &= [1, x_1, x_2, {x_1}^2, {x_2}^2, x_1 x_2], \quad \vec{\omega} = [\omega_0, \ldots, \omega_5] \\ k = 3 : \vec{x} &= [1, x_1, x_2, {x_1}^2, {x_2}^2, x_1 x_2, {x_1}^3, {x_2}^3, x_1 {x_2}^2, \\ &\quad {x_1}^2 x_2], \quad \vec{\omega} = [\omega_0, \ldots, \omega_9]. \end{aligned} \tag{1}$$

The estimation of the parameters $\vec{\omega}$ of the polynomial is performed by linear least squares QR factorization [20]. We limit our polynomial approximations to the third order, $k = 3$, to prevent overfitting the data.

The second step is to select the ground-truth base layer for each part. For each part, there are four possible ground-truth base layers, and we let the annotators select the ground-truth base layer by choosing one from the four layers.

After the previous two steps, we get the ground-truth base layer using the polynomial model's separation results and the annotation of the annotators on each part. It is possible that none of the possible results obtained by the polynomial model is correct for the base layer. Thus, for an image, if the base layer of even one part cannot be described by any of the polynomial results, this image is rejected. In this way, we get a subset of the images from the whole data sets of the Fashionista and the Pascal Part UCLA, for which the base layer can be described by some order of the polynomial model. In total, we select about 1000 and 250 images in the Pascal Part UCLA and the Fashionista data sets, respectively.

In our labeling, 15 annotators in total performed the labeling separately. So, for each region of the images, we have 15 annotations. If seven or more of the 15 annotations choose "outlier region," we will see this region as an outlier to be modeled by the polynomial model and do not select the image of the object into our final data sets. Otherwise, the order of the polynomial for this region is voted by the 15 annotations. And the base layer of the region is reconstructed by the polynomial model.

IV. Evaluation

A. Ground-Truth Data Sets

The ground-truth data sets that we use are the subsets of images from the Fashionista [18] and the Pascal Part UCLA data sets [19], as described in Section III. For simplification, in this letter, we still call the subset of images the Fashionista and the Pascal Part UCLA data sets, respectively. See the examples of both data sets in Fig. 5.

B. Separation Methods

The separation methods we use in the evaluation include the adaptive manifold (AM) filter [8], the domain transform (DT) filter [7], the L0 smooth (L0) filter [1], the rolling guidance (RG) filter [9], the bilateral (BL) filter [10], the guided (GD) filter [11], and the Gaussian (GS) filter. The GS filter is a linear filter, and the smoothing considers only the distance between neighboring pixels. The other filters are edge-preserving filters.

C. Error Metric

A direct way for evaluation is to compute the mean squared error (mse) between ground-truth base-detail layers and estimated base-detail layers. The mse is defined as

$$\mathrm{MSE}(J_1, J_2) = \frac{1}{\sum_{i,c} 1} \sum_{i,c} (J_1(i, c) - J_2(i, c))^2$$

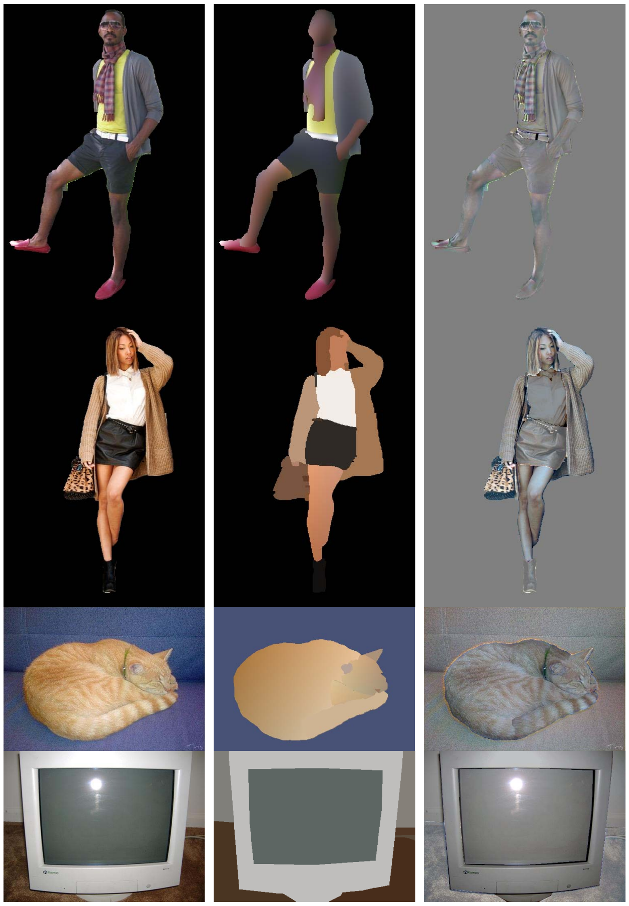

Fig. 5. Examples of the images in the Fashionista (top two) and the Pascal Part UCLA (bottom two) base-detail data sets. From left to right: original image, base, and detail.

where J_1 and J_2 are two images, i is the pixel position, and c is the color channel in the RGB space.

However, we found that the same amount of error will cause very different mse values for well-exposed images and low lighting images. For example, the intensities of a pixel in a well-exposed image and low-lighting image are 200 and 20, respectively. If the errors are the same, for example 10%, the mse values will be very different (400 and 4, respectively). The reason is that for well-exposed images, because the RGB intensities of pixels are high, small errors will lead to large mse values. So directly using mse will lead to bias to the evaluation results, and the errors of well-exposed images will have more weights. To reduce the bias, we proposed the relative mse (RMSE) as the error metric between the ground-truth base-detail layers and the estimated base-detail layers. The RMSE is defined by

$$\begin{aligned}\mathrm{RMSE}(B_{\mathrm{GT}}, D_{\mathrm{GT}}, B_E, D_E)\\ = \frac{1}{2}\left(\frac{\mathrm{MSE}(B_{\mathrm{GT}}, B_E)}{\mathrm{MSE}(B_{\mathrm{GT}}, 0)} + \frac{\mathrm{MSE}(D_{\mathrm{GT}}, D_E)}{\mathrm{MSE}(D_{\mathrm{GT}}, 0)}\right) \quad (2)\end{aligned}$$

where B_{GT} is the ground-truth base layer, D_{GT} is the ground-truth detail layer, B_E is the estimated base layer, and D_E is the estimated detail layer. Because: 1) RMSE considers errors of both the detail and the base layers and 2) for each layer, it measures the relative error, i.e., the ratio of $\mathrm{mse}(\mathrm{GT}, E)$ and $\mathrm{mse}(\mathrm{GT}, 0)$, it reduces the bias between low-lighting images and well-exposed images. According to the definition of RMSE, if the RMSE value is lower, the estimation of base and detail layers are more accurate.

TABLE I

PARAMETERS SETTINGS OF DIFFERENT FILTERS FOR THE FASHIONISTA AND THE PASCAL PART UCLA DATA SETS. LEFT: PARAMETERS RANGE OF THE FILTERS. RIGHT: OPTIMAL PARAMETERS OF THE METHODS FOR THE DATA SETS

	Ranges of parameters θ_i		Optimal (θ_1, θ_2)	
Alg.	θ_1	θ_2	Fashionista	Pascal
AM	2,4,8,16,32,64	0.1,0.2,0.4,0.8,1,2	(8,0.8)	(32,0.8)
DT	4,8,16,32,64,128	0.1,0.2,0.4,0.8,1,2	(16,2)	(64,2)
L0	0.005,0.01,0.02,0.04,0.08,0.1,0.2,0.4,0.8,1		0.08	0.1
RG	2,4,8,16	0.04,0.08,0.1,0.2,0.4	(8,0.1)	(16,0.2)
BL	2,4,8,16,32,64	0.1,0.2,0.4,1,2,4	(8,0.4)	(32,0.4)
GD	4,8,16,32,64,128	0.01,0.04,0.1,0.2,0.4	(16,0.1)	(64,0.1)
GS	4,8,16,32,64,128	1,2,4,8,16,32	(32,4)	(64,4)

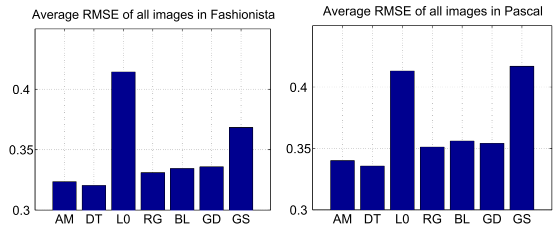

Fig. 6. Quantitative comparison of the seven separation algorithms on the data sets (from left to right): average RMSE on the Fashionista and the Pascal Part UCLA (over all images in the data set).

D. Algorithms and Parameter Settings

For an input image, we use different algorithms to smooth it to obtain the base layers. Then, we compute the RMSE between the filtered result and the ground-truth data. The seven separation methods shown in Table I have parameters to control the smoothing. In general, high values of the parameters tend to mean coarse level smoothing. Here, we select the best parameters for each filter to enable a fair comparison among them. We use the same parameter value for the whole data set (one parameter for each data set). The parameters range of the filters and the optimal parameters are shown in Table I. For GS and GD, θ_1 and θ_2 denote window size and σ variance, respectively. For BL, AM, DT, and RG, θ_1 and θ_2 denote σ spatial and σ range, respectively. For L0, θ_1 denotes λ. The parameters for the Fashionista and the Pascal Part UCLA data sets are different because, as described in Section III-A, these two data sets used different strategies for part annotation. The segmented parts of the Pascal Part UCLA data set are usually coarser than those of the Fashionista data set. As a result, the optimal parameters for the Pascal Part UCLA data set are usually larger than those for the Fashionista data set. The results are shown in Fig. 6. Some example results of each method can be seen in Figs. 3 and 7.

E. Analysis

We can see that most of the edge-preserving filters perform better than the GS filter (we consider the GS filter to be the classical baseline). It is because the GS filter does not preserve edges, and the high parameters (e.g., the variance) lead to the smoothing results of each part affected heavily by neighboring parts. Most of the other filters preserve edges better than the GS filter and so they have better performance. The BL filter is edge-preserving and consistent with standard intuition and so it performs better than the GS filter. The AM filter and DT filter have better performance than the BL filter on average because they are flexible and have more potential to perform well in our data sets if the parameters are selected carefully. The RG filter and GD filter also have good performances in the experiments. The L0 filter makes use of gradient to separate base

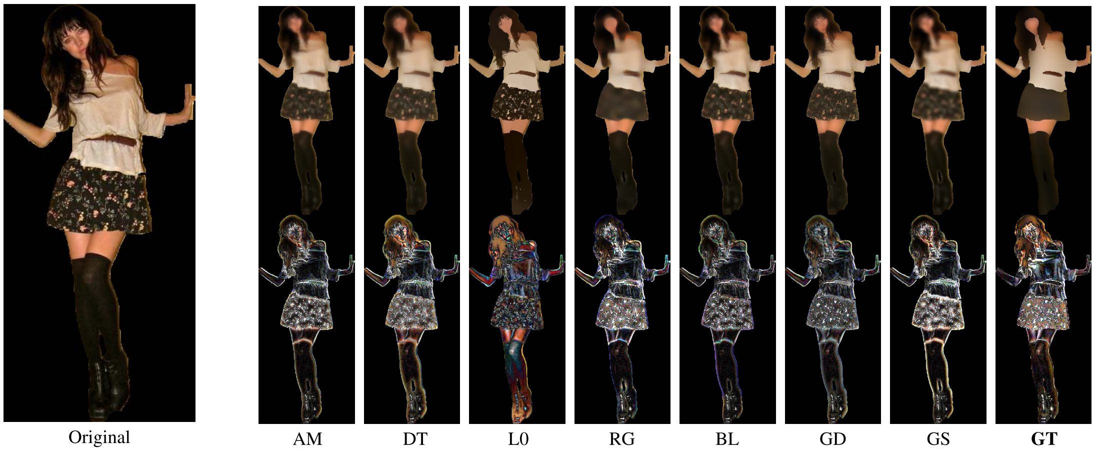

Fig. 7. Results of the base-detail separation methods tested on an example image of the Fashionista dataset. Top: base. Bottom: detail. Last column (GT) is the proposed ground-truth.

and detail. However, in our data sets, the parts are segmented semantically, and areas with large gradient do not always mean the edges of parts. This results in poor performance.

V. Conclusion

Quantitative evaluations are fundamental to the advances of any research field. The part-level base-detail ground-truth data sets we provide here are a necessary starting point for extensive quantitative comparisons of the base-detail separation algorithms at the part level. We argue that, ideally, base-detail annotation should be hierarchical, and we proposed an intermediate solution, which is practical (i.e., ready for use) now and extensible in the future.

Acknowledgment

The authors would like to thank the anonymous reviewers for their help in improving this letter.

References

[1] L. Xu, C. Lu, Y. Xu, and J. Jia, “Image smoothing via L_0 gradient minimization,” *ACM Trans. Graph.*, vol. 30, no. 6, Dec. 2011, Art. no. 174.

[2] L. Yuan and J. Sun, “Automatic exposure correction of consumer photographs,” in *Proc. 12th ECCV*, 2012, pp. 771–785.

[3] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, “A naturalistic open source movie for optical flow evaluation,” in *Proc. 12th ECCV*, 2012, pp. 611–625.

[4] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.

[5] L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories,” *Comput. Vis. Image Understand.*, vol. 106, no. 1, pp. 59–70, Jan. 2007.

[6] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[7] E. S. L. Gastal and M. M. Oliveira, “Domain transform for edge-aware image and video processing,” *ACM Trans. Graph.*, vol. 30, no. 4, p. 69, 2011.

[8] E. S. L. Gastal and M. M. Oliveira, “Adaptive manifolds for real-time high-dimensional filtering,” *ACM Trans. Graph.*, vol. 31, no. 4, Jul. 2012, Art. no. 33.

[9] Q. Zhang, X. Shen, L. Xu, and J. Jia, “Rolling guidance filter,” in *Proc. 13th ECCV*, 2014, pp. 815–830.

[10] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, “Bilateral filtering: Theory and applications,” *Found. Trends Comput. Graph. Vis.*, vol. 4, no. 1, pp. 1–73, 2009.

[11] K. He, J. Sun, and X. Tang, “Guided image filtering,” in *Proc. 11th ECCV*, 2010, pp. 1–14.

[12] T. Brox and D. Cremers, “Iterated nonlocal means for texture restoration,” in *Proc. Int. Conf. Scale Space Variat. Methods Comput. Vis.*, 2007, pp. 13–24.

[13] P. Bhat, C. L. Zitnick, M. Cohen, and B. Curless, “GradientShop: A gradient-domain optimization framework for image and video filtering,” *ACM Trans. Graph.*, vol. 29, no. 2, Mar. 2010, Art. no. 10.

[14] F. Durand and J. Dorsey, “Fast bilateral filtering for the display of high-dynamic-range images,” *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, Jul. 2002.

[15] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand, “Fast local Laplacian filters: Theory and applications,” *ACM Trans. Graph.*, vol. 33, no. 5, Aug. 2014, Art. no. 167.

[16] S. Bae, S. Paris, and F. Durand, “Two-scale tone management for photographic look,” *ACM Trans. Graph.*, vol. 25, no. 3, pp. 637–645, 2006.

[17] B. Bonev and A. L. Yuille, “A fast and simple algorithm for producing candidate regions,” in *Proc. 13th ECCV*, 2014, pp. 535–549.

[18] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg, “Parsing clothing in fashion photographs,” in *Proc. IEEE Conf. CVPR*, Jun. 2012, pp. 3570–3577.

[19] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille, “Detect what you can: Detecting and representing objects using holistic models and body parts,” in *Proc. IEEE Conf. CVPR*, Jun. 2014, pp. 1971–1978.

[20] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2012.