

Focus-and-Context Skeleton-Based Image Simplification Using Saliency Maps

Jieying Wang¹^a, Leonardo de Melo Joao²^b, Alexandre Falcão²^c,
Jiří Kosinka¹^d, and Alexandru Telea³^e

¹Bernoulli Institute, University of Groningen, 9747 AG Groningen, The Netherlands

²Department of Information Systems, Institute of Computing, University of Campinas, São Paulo CEP 13083-852, Brazil

³Department of Information and Computing Sciences, Utrecht University, 3584 CC Utrecht, The Netherlands
jieying.wang@rug.nl, leomelo168@gmail.com, afalcao@ic.unicamp.br, j.kosinka@rug.nl, a.c.telea@uu.nl

Keywords: Medial axis, Dense skeleton, Image simplification, Saliency map

Abstract: Medial descriptors offer a promising way for representing, simplifying, manipulating, and compressing images. However, to date, these have been applied in a global manner that is oblivious to salient features. In this paper, we adapt medial descriptors to use the information provided by saliency maps to selectively simplify and encode an image while preserving its salient regions. This allows us to improve the trade-off between compression ratio and image quality as compared to the standard dense-skeleton method while keeping perceptually salient features, in a focus-and-context manner. We show how our method can be combined with JPEG to increase overall compression rates at the cost of a slightly lower image quality. We demonstrate our method on a benchmark composed of a broad set of images.


1 INTRODUCTION


Images are one of the most widely present data types in many application areas in science, engineering, but also end-user applications. Image compression and simplification are two closely related techniques in the toolset of the imaging practitioner: *Compression* creates images of a smaller file size for archiving, transmission, and rendering purposes; lossy compression achieves this by removing certain image details (or parts thereof), though typically not in an explicitly user-controlled manner. *Simplification* creates images which keep visual structures of interest to the use-case at hand, and remove the other, less important, structures, to ease further analysis and processing of such images; simplification also achieves image-size reduction, although as a by-product rather than a key goal.


Recently, Dense Medial Descriptors (DMDs) have been proposed as a new way to perform image compression (Wang et al., 2020b). DMDs model an im-


age as a collection of luminance threshold-sets, or layers, each layer being encoded by its medial axis transform (MAT). DMDs create a simplified versions of an image by suitably selecting a subset of its layers and storing simplified (pruned) versions of their MATs. Qualitative and quantitative evaluation has shown that DMDs deliver good compression ratios while preserving image quality. As such, DMDs can be an interesting and promising option for lossy image encoding. However, DMDs offer so far only a *global* way to simplify an image, which does not always lead to optimal quality, as certain details deemed important by the user may be simplified away together with less important details.


In this paper, we extend the DMD method (Wang et al., 2020b) with a so-called *spatial saliency map* that models the importance of various areas in an image. We use this map to control image simplification, thereby enabling finer-grained *spatial* control of the simplification. This makes DMDs suitable for applications such as focus-and-context compression. We propose several metrics to gauge the effectiveness of our method and the trade-off between image size and perceptual similarity. We evaluate these metrics on a collection of real-world images to illustrate the advantages of our extended method.

^a <https://orcid.org/0000-0002-0085-3551>

^b <https://orcid.org/0000-0003-4625-7840>

^c <https://orcid.org/0000-0002-2914-5380>

^d <https://orcid.org/0000-0002-8859-2586>

^e <https://orcid.org/0000-0003-0750-0502>

The remainder of the paper is organized as follows. We start with an introduction of the background in Section 2, including dense medial descriptors, saliency maps, and image quality metrics. Section 3 describes our proposed modifications to DMD method to include saliency-aware image simplification. Section 4 details the obtained results. Section 5 discusses our results. Finally, Section 6 concludes the paper.

2 BACKGROUND

We start by outlining related work regarding (dense) medial descriptors, saliency maps, and image quality metrics.

2.1 Dense Medial Descriptors

We first briefly describe the DMD method (Fig. 1). For full details, we refer to (Wang et al., 2020b). Let $I : \mathbb{R}^2 \rightarrow [0, 255]$ be an 8-bit grayscale image. All results next apply to color images too by considering each of their three channels in *e.g.* YUV color space. I is reduced to $n = 256$ threshold sets or layers $T_i = \{\mathbf{x} \in \mathbb{R}^2 \mid I(\mathbf{x}) \geq i\}$, $0 \leq i < n$. From these, a subset of $L < n$ layers is kept, by removing layers which are very similar to each other, thus contribute little to describing I . Next, islands (connected components in the foreground T_i or background \bar{T}_i) that are smaller than a user-given threshold ϵ , thus contribute little to the image, are removed. Next, a binary skeleton S_i is extracted from each layer T_i . Such skeletons contain spurious branches caused by small perturbations along the boundary ∂T_i of T_i . These can be eliminated by regularization (Telea and van Wijk, 2002; Costa and Cesar, 2000; Falcão et al., 2004). DMD uses the so-called salient-skeleton regularization metric (Telea, 2012) which keeps perceptually sharp corners of ∂T_i but removes small-scale wiggles along ∂T_i , based on a user parameter $\sigma_0 > 0$. The parameter σ_0 has a geometric meaning: Setting $\sigma_0 = 0.1$ means removing all wiggles smaller than 10% of the local object thickness (Telea, 2012). A simplified version of the image I is finally reconstructed from the regularized skeletons \tilde{S}_i and their distance transforms, *i.e.*, the simplified medial axis transforms (MATs) of the selected layers T_i . The parameter σ_0 controls the scale of image details to be removed, thereby enabling applications such as image segmentation (Koehoorn et al., 2015; Sobiecki et al., 2015; Koehoorn et al., 2016) and nonphotorealistic image rendering (Zwan et al., 2013).

However, DMD can only simplify an image *globally*. High simplification will easily remove small, but visually important, details. Conversely, low simplification will allocate storage to unimportant image areas (poor compression). In many cases, users may want to keep (or remove) same-scale details based on the *context* these appear in. Our method, described in Sec. 3, adapts DMD precisely in this direction, that is, to use context information to drive the simplification.

2.2 Saliency Maps

Saliency maps $\mu : \mathbb{R}^2 \rightarrow [0, 1]$ encode how important each image pixel is for a given task or perceptual standpoint (0 being totally unimportant and 1 being of maximal importance). Such maps have been used for image quality assessment (Liu and Heynderickx, 2011), content-based image retrieval (Chen et al., 2009), context-aware image resizing (Goferman et al., 2011), image compression (Andrushia and Thangarajan, 2018; Zünd et al., 2013), and saliency-based gaze tracking (Cazzato et al., 2020). They can be computed by several techniques, as follows.

Supervised methods, especially those using deep learning, typically outperform unsupervised methods when enough training images are used (Borji et al., 2015). Saliency estimators are commonly evaluated using image segmentation metrics on a thresholded saliency map, yielding binary saliency with possibly unnatural values for some regions of a salient object. Binary saliency maps are useful for segmentation; smoother (non-binary) maps allow a more continuous selection of important *vs* less important image areas, as needed in our case (see next Sec. 3).

Unsupervised methods propose heuristics to model what makes objects salient in a scene. Most methods start by finding image regions (*e.g.* superpixels) with high color contrast relative to neighbors (Jiang et al., 2013; Li et al., 2013; Zhang et al., 2018). Besides contrast, objects in focus (Jiang et al., 2013), near the image center (Cheng et al., 2014), or having red and yellow tones (important for the human visual system) (Peng et al., 2016), are likely salient. Since most image boundaries are background in natural images, regions similar to the boundary will have low saliency (Cheng et al., 2014; Zhang et al., 2018; Li et al., 2013; Jiang et al., 2013). Unsupervised saliency estimators combine several such assumptions. In our work, we apply such an unsupervised bottom-up saliency estimation algorithm, namely DSR (Li et al., 2013), which provides reliable saliency maps without requiring parameter tuning, and in a short amount of time. However, any other saliency map methods can be directly used instead, in-

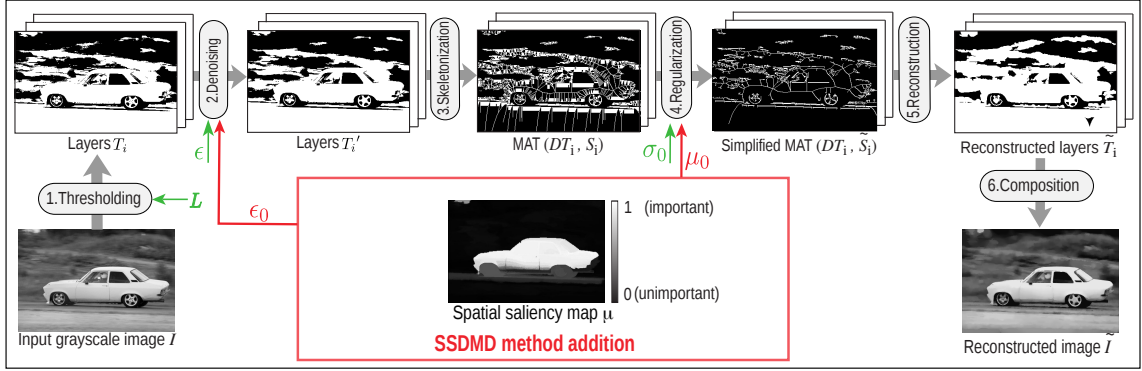


Figure 1: Dense medial descriptor (DMD) pipeline with free parameters in green. Red: Elements added by our SSDMD method.

cluding manually designed maps, as long as users find the produced maps suitable for their tasks at hand.

2.3 Image Quality Metrics

A quality metric $Q(I, \tilde{I}) \in \mathbb{R}^+$ measures how perceptually close an image I is to its representation \tilde{I} . Such metrics include the mean squared error (MSE) and peak signal-to-noise ratio (PSNR). While simple to compute and with clear physical meanings, these do not match well perceived visual quality (Wang and Bovik, 2009; Zhang et al., 2011; Zhang et al., 2012). We illustrate this in the supplementary material (Wang et al., 2020a). The structural similarity (SSIM) index (Wang et al., 2004) alleviates this by measuring, pixel-wise, how similar two images are by considering human perception. Mean SSIM (MSSIM) aggregates SSIM to a scalar value by averaging over all image pixels. MSSIM was extended to three-component SSIM (3-SSIM) by using non-uniform weights for the SSIM map over three region types: edges, texture, and smooth areas (Li and Bovik, 2010). Multiscale SSIM (MS-SSIM) (Wang et al., 2003) is an advanced top-down interpretation of how the human visual system interprets images that considers variations of image resolution and viewing conditions. Comprehensive evaluations (Sheikh et al., 2006; Ponomarenko et al., 2009) have demonstrated that SSIM and MS-SSIM can offer statistically much better performance in assessing image quality than other quality metrics. Moreover, as MS-SSIM outperforms the best single-scale SSIM model (Wang et al., 2003), we consider it next in our work.

3 PROPOSED METHOD

As stated in Sec. 2.1, an important limitation of DMD is that it simplifies an image *globally*. There-

fore, we improve DMD by considering *spatially-dependent* simplification of image foreground and background. We call our method Spatial Saliency DMD (SSDMD for short). Fig. 1 (red) shows the steps that SSDMD adds to DMD. These steps are described next.

3.1 Salient Islands Detection

As explained in Sec. 2, DMD removes islands smaller than a *global* value of ϵ area units. This removes not only noise but also small important features (e.g. the animal eyes in Fig. 3 a1–c1). To address this, we compute a saliency-aware metric $C_i^\mu = \sum_{\mathbf{x} \in C_i} \mu(\mathbf{x})$, where C_i is the i th connected component, and next remove only islands where C_i^μ is below a user-given threshold ϵ_0 . This keeps small-size, but salient, details, in the compressed image.

3.2 Saliency-based Skeletons

We further simplify the regularized skeletons \tilde{S}_i by removing pixels whose saliency μ is below a user-given threshold μ_0 , resulting in saliency-aware skeletons $\tilde{S}_i^\mu = \{\mathbf{x} \in \tilde{S}_i | \mu(\mathbf{x}) > \mu_0\}$. The threshold μ_0 controls the amount of the non-salient areas. To avoid low-saliency areas (with saliency μ that are below the global threshold μ_0) being completely removed, resulting in poor image quality, we reserve one layer every m layers for these areas. The skeletons \tilde{S}_i to be reconstructed are then computed using the piecewise formulation

$$\tilde{S}_i = \begin{cases} \tilde{S}_i, & \text{if } i \bmod m = 0, \\ \tilde{S}_i^\mu, & \text{otherwise.} \end{cases}$$

The parameter m controls how smooth color or brightness gradients will be in the non-salient areas; smaller m values yield smoother gradients. Since only several

layers are reserved in non-salient areas, an intensity-banding effect can occur. To solve this, we apply a smooth distance-based interpolation between two consecutive selected layers T_i and T_{i+1} (Zwan et al., 2013). In detail, for a pixel \mathbf{x} located between the boundaries ∂T_i and ∂T_{i+1} , we interpolate its value $I(\mathbf{x})$ (in all three channels independently) from the corresponding values I_i and I_{i+1} as

$$I(\mathbf{x}) = \frac{1}{2} \left[\min\left(\frac{DT_i}{DT_{i+1}}, 1\right) I_i + \max\left(1 - \frac{DT_i}{DT_{i+1}}, 0\right) I_{i+1} \right],$$

where DT_i is the distance transform of layer T_i evaluated at location \mathbf{x} .

3.3 Saliency-aware Quality Metric

While MS-SSIM models human perception well (Sec. 2.3), it treats focus (high $\mu(\mathbf{x})$) and context (low $\mu(\mathbf{x})$) areas identically. Figure 2 shows this: Image (a) shows the DMD compression of a car image. Image (b) shows the SSIM map, *i.e.*, the per-pixel structural similarity between the original image and its DMD compression, in which darker pixels indicate lower similarity. Image (a) shows some artifacts on the car roof, also visible as dark regions in the SSIM map (b). Image (c) shows the SSDMD compression of the same image, with strong background simplification and high detail retention in the focus (car) area. The car-roof compression artifacts are removed, so (c) is a better representation than (a) of the original image. However, the MS-SSIM score of (c) is much lower than for DMD compression (0.9088 *vs* 0.9527). The large dark areas in the background of the SSIM map (d) explain this: While our saliency map μ clearly says that background is unimportant, MS-SSIM considers it *equally* important as foreground, which is counter-intuitive.

Given the above, saliency data should be (visually) considered in the quality metric so that the latter is more consistent with the human visual system. This is also reflected by saliency-based objective metrics reported in the literature (Le Callet and Niebur, 2013; Engelke and Le Callet, 2015; Liu and Heynderickx, 2011; Liu et al., 2013; Alaei et al., 2017). In these designs, a visual saliency map is integrated into the quality metric as a weighting map, which improves image quality prediction performance. We follow the same idea, by integrating the spatial saliency map into the MS-SSIM (Wang et al., 2003) pooling function, as follows. Take the MS-SSIM metric for a reference image I and a distorted image \tilde{I}

$$Q(I, \tilde{I}) = [\text{SSIM}(I, \tilde{I})]^{\beta_M} \prod_{j=1}^{M-1} [c_j(I, \tilde{I})]^{\beta_j}, \quad (1)$$

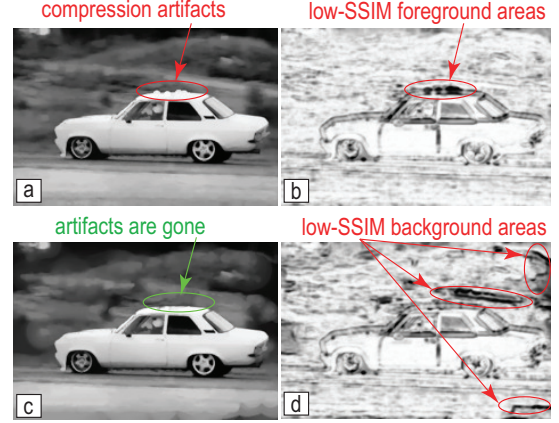


Figure 2: DMD compression has artifacts (a) found as low-SSIM regions (b). SSDMD (c) removes these but finds subtle background differences as important for quality (d).

where c_j is the contrast map $c(I, \tilde{I})$ iteratively down-sampled by a factor of 2 on scale $1 \leq j \leq M$ and $\text{SSIM}(I, \tilde{I})$ is the structural similarity of I and \tilde{I} on scale M (Wang et al., 2004). The exponent β_j models the relative importance of different scales. We weigh Q by the saliency map μ , yielding the saliency-aware quality metric

$$Q^\mu = \left[\frac{\sum_{\mathbf{x} \in I} \mu(\mathbf{x}) \text{SSIM}(\mathbf{x})}{\sum_{\mathbf{x} \in I} \mu(\mathbf{x})} \right]^{\beta_M} \prod_{j=1}^{M-1} \left[\frac{\sum_{\mathbf{x} \in I} \mu_j(\mathbf{x}) c_j(\mathbf{x})}{\sum_{\mathbf{x} \in I} \mu_j(\mathbf{x})} \right]^{\beta_j}, \quad (2)$$

where μ_j is the saliency map at scale j . For notation brevity, we omitted the arguments I and \tilde{I} in Eqn. 2. Using Q^μ instead of Q allows in-focus values (high $\mu(\mathbf{x})$) to contribute more to similarity than context values (low $\mu(\mathbf{x})$), in line with our goal of spatially-controlled simplification.

4 RESULTS

The proposed SSDMD method described in Sec. 3 adapts the original DMD pipeline by using the spatial saliency information. We next demonstrate SSDMD, and discuss its properties, on several images. In the following, we define the compression ratio of an image as $CR = |I|/|MAT(\tilde{I})|$, *i.e.*, the size (in bytes) of the original I divided by the size (in bytes) of the MATs of the L selected layers used to encode \tilde{I} . The latter includes the size of the encoded file that needs to be stored to reconstruct the original image using the (SS)DMD method.

Increasing compression while retaining highlights

Figure 3 shows the simplification of three bird images by DMD (a1–c1) and SSDMD (a2–c2). To test our

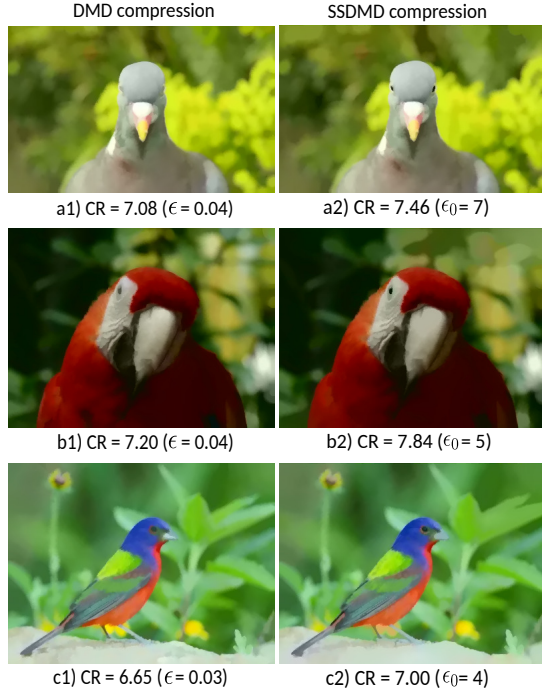


Figure 3: Comparison of DMD (a1–c1) with SSDMD (a2–c2). The compression ratio CR is indicated for each image.

new saliency-aware metric C_i^μ (Sec. 3.1), we keep all parameter settings of DMD and SSDMD the same, and only vary the island detection parameter ϵ and ϵ_0 for DMD and SSDMD separately. The identical parameters are set to empirically-determined values (Wang et al., 2020b), *i.e.*, $L = 30$ and $\sigma_0 = 0.1$. Compared to DMD, SSDMD preserves the birds’ eyes while simplifying the background more, which allows it to achieve higher compression ratios while keeping perceptually salient features.

Q^μ achieves higher correlation with human perception Figure 4 shows DMD (a1–c1) and SSDMD (a2–c2) applied to three focus-and-context images. For each image, we indicate the standard MS-SSIM quality Q , spatial-saliency-aware quality Q^μ , and compression ratio CR. The Q values for SSDMD are lower than those for DMD, which suggests that SSDMD has a poorer quality than DMD. Yet, we see that SSDMD produces images that are visually almost identical to DMD, in line with the almost identical Q^μ values for SSDMD and DMD. Thus, we argue that Q^μ is a better quality measure for focus-and-context simplification than Q . Also, we see that, while Q^μ stays almost identical, SSDMD compresses better than DMD (CR values on average 24.6% higher).

Increasing compression and/or quality Figure 5 extends this insight to 150 images, selected randomly



Figure 4: Comparison of DMD (a1–c1) with SSDMD (a2–c2) for three focus-and-context images. For each image, we show the standard MS-SSIM quality Q , spatial-saliency-aware MS-SSIM Q^μ , and compression ratio CR.

from the MSRA10K (Cheng, 2014), SOD (Movahedi and Elder, 2010), and ECSSD (Shi et al., 2016) benchmarks. Hollow dots in Fig. 5 are DMD compression results, and filled dots are SSDMD results. One dot represents the average Q^μ and CR for a specific parameter-setting over *all images* in the benchmark. Same-kind dots show $2 \cdot 3 \cdot 3 = 18$ different settings of the parameters L , ϵ , and σ_0 (actual values shown in Fig. 5). To find these, we first evaluated Q^μ and CR by grid search over the full allowable ranges of L , ϵ , and σ_0 , and then found subranges where both Q^μ and CR yielded high values. Next, we took a few samples within these subranges, leading to the values shown in the figure. Finally, we set threshold $\mu_0 = 0.01$, *i.e.*, keeping all but the least salient parts of the image; recall that $\mu(\mathbf{x}) \in [0, 1]$.

As explained in Sec. 2.1, for color images, (SS)DMD is applied to the individual channels of these, following representations in various color spaces. In contrast to (Wang et al., 2020b), which uses the RGB color space, we choose to use YUV (more precisely, YCbCr) in all the (SS)DMD experiments, for two reasons. First, YUV was shown to give better subjective quality than RGB due to its perceptual similarities to human vision (Podpora et al., 2014; Podpora, 2009). Secondly, since the human eye is less sensitive to the chrominance components Cb (blue projection) and Cr (red projection), strongly compressing these components achieves a higher com-

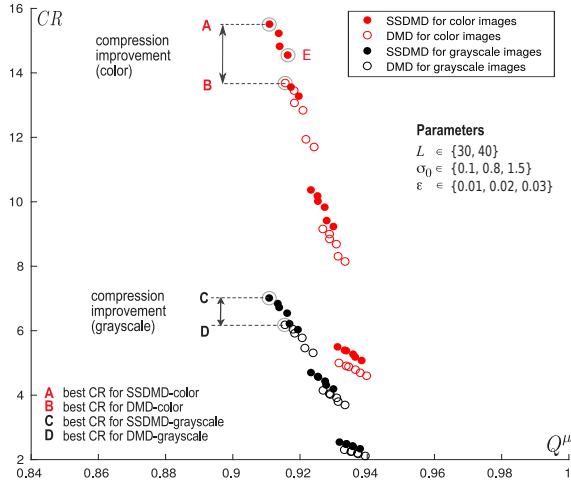


Figure 5: Average quality Q^μ vs compression ratio CR for 150 images for SSDMD and DMD.

pression ratio while keeping quality high (Nobuhara and Hirota, 2004). We see this also in Fig. 5: The SSDMD compression ratio (CR) of color images (red dots) is more than twice that of the grayscale images (black dots) on average, and nearly always higher than the CR of the same images computed by DMD (red circles), while having the same quality. We also observe that for both color and grayscale images, the best CR values we obtain with SSDMD (points A, C) is about 14% higher than the best CR produced by DMD (points B, D). Hence, SSDMD improves compression as compared with DMD, with, as visible in Fig. 5, only a very slight decrease in quality Q^μ . In particular, point E shows a run of SSDMD that improves on *both* compression (CR) and quality (Q^μ) as compared to the highest-compression run of DMD (point B). Figure 6 further explores this insight for six real-world images (plant, animal, natural scene, people, and man-made structure) from the MSRA10K, SOD, and ECSSD benchmarks. We show both color versions and their grayscale counterparts compressed by DMD and SSDMD, and their corresponding CR and Q^μ values. We also show their saliency maps μ on top to illustrate what is considered focus and context. Both images and values in Fig. 6 show that the SSDMD method *increases the compression ratio while maintaining perceived quality*.

Progressively simplification effect of μ_0 As already discussed, Fig. 5 compares 18 different settings of the parameters L , ϵ , and σ_0 for both DMD and SSDMD, for a fixed value $\mu_0 = 0.01$. This was done to ease the interpretation of the respective scatterplots, as using multiple μ_0 values in the same figure would have been hard to read. However, the parameter μ_0 does affect the CR vs Q^μ trade-off, effec-

tively allowing the user to specify how strongly s/he wants to simplify the image (increase CR) by trading off a certain quality amount (decrease Q^μ). Figure 7 gives insight into this, showing three images (flower in the saliency focus in all cases) for three settings $\mu_0 \in \{0.04, 0.08, 0.12\}$. The setting $\mu_0 = 0$ corresponds to DMD. All other parameters are fixed to default values $L = 50$, $\epsilon = 0.01$, $\sigma_0 = 0.5$, and $m = 8$. Compared with DMD, the background areas of the SSDMD images are gradually simplified as μ_0 increases; however, the flower is not changed, as it is in a high-saliency area. The CR and Q^μ values shown below the images show that increasing μ_0 greatly improves the compression ratio of SSDMD while quality is only slightly reduced.

JPEG preprocessor A final interesting use-case is to *combine* SSDMD’s simplification ability with a generic image compressor. For this, we ran SSDMD as a ‘preprocessor’ and subsequently compressed its result with standard JPEG. Figure 8 shows the results of plain JPEG compression at 20% quality setting and SSDMD+JPEG for the same quality setting for three images. Values in green are the CR of SSDMD+JPEG divided by plain JPEG’s CR , i.e., the compression *gain* when using SSDMD as preprocessor for JPEG. This gain is 15%, 12% and 21% for the church, car, and spectacles image, respectively. For these images, the results using SSDMD+JPEG are visually almost identical in the focus areas (church building, car shape, and spectacles shape). Of course, in the context area (sky around church, scenery around car, book around spectacles) some differences are visible. This is expected — and intended — since, as explained, SSDMD aims to keep details in the focus area while simplifying them away in the context. Table 1 extends these insights by listing results under additional JPEG quality setting values for these

Table 1: Performance of plain JPEG and SSDMD + JPEG under different quality settings for the images in Fig. 8.

Images	Quality Settings (%)	Plain JPEG (CR/Q^μ)	SSDMD + JPEG (CR (gain)/ Q^μ)
Church	40	61.9/0.991	72.1 (1.16)/0.955
	60	47.1/0.994	55.6 (1.18)/0.958
	80	30.7/0.997	37.5 (1.22)/0.960
	100	4.3/1.0	6.7 (1.55)/0.961
Car	40	37.8/0.994	44.7 (1.18)/0.942
	60	28.5/0.996	34.0 (1.19)/0.943
	80	19.2/0.998	22.9 (1.19)/0.944
	100	3.5/1.0	4.3 (1.23)/0.944
Spectacles	40	38.0/0.995	46.7 (1.23)/0.957
	60	29.4/0.997	36.4 (1.24)/0.958
	80	20.0/0.999	25.2 (1.26)/0.959
	100	5.2/1.0	6.0 (1.15)/0.960

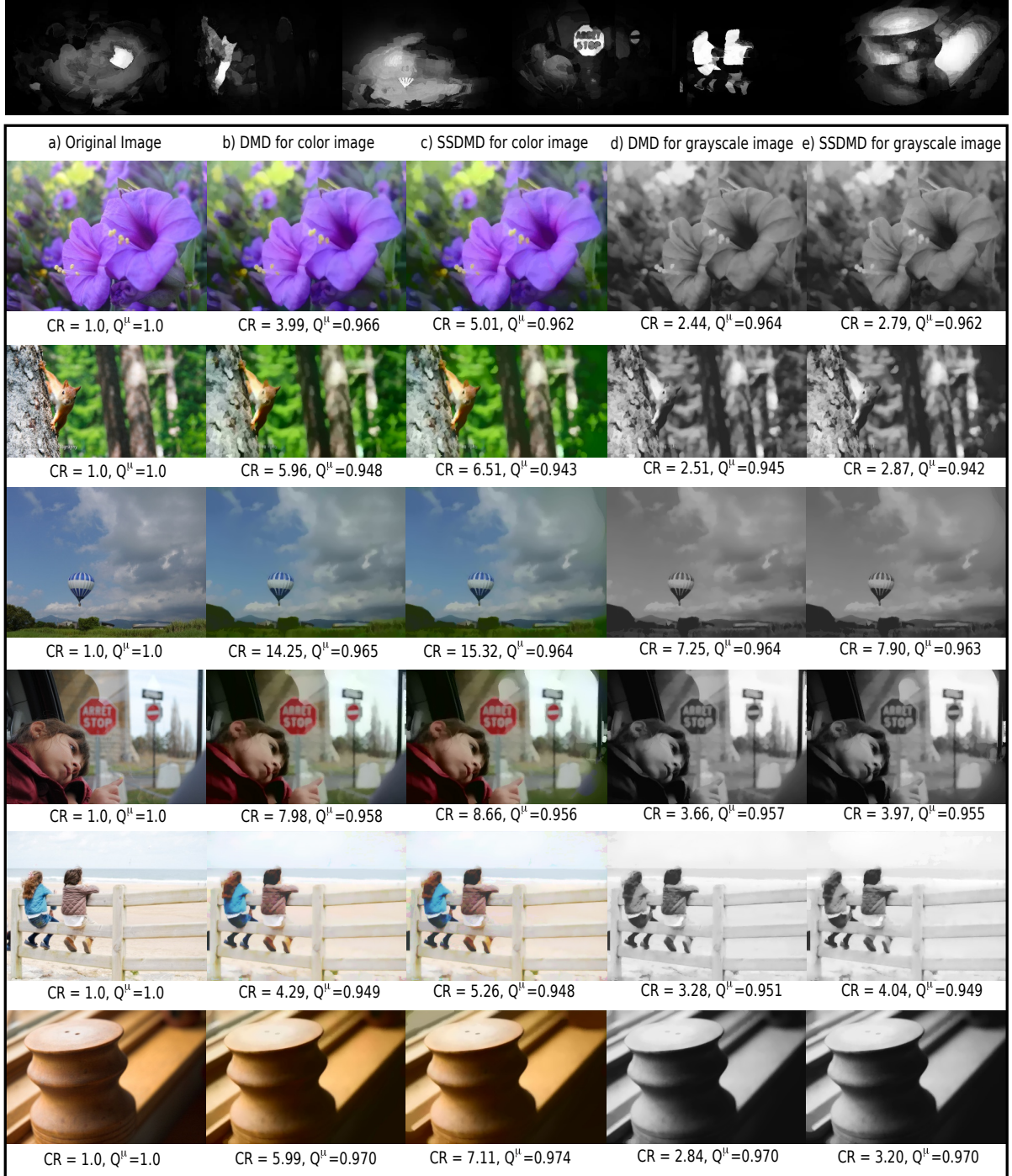


Figure 6: Comparison of the DMD with the SSDMD method for color and grayscale versions of six input images. The top row shows the spatial saliency maps of each input image. For each image, we show the compression ratio CR and quality score Q^u .

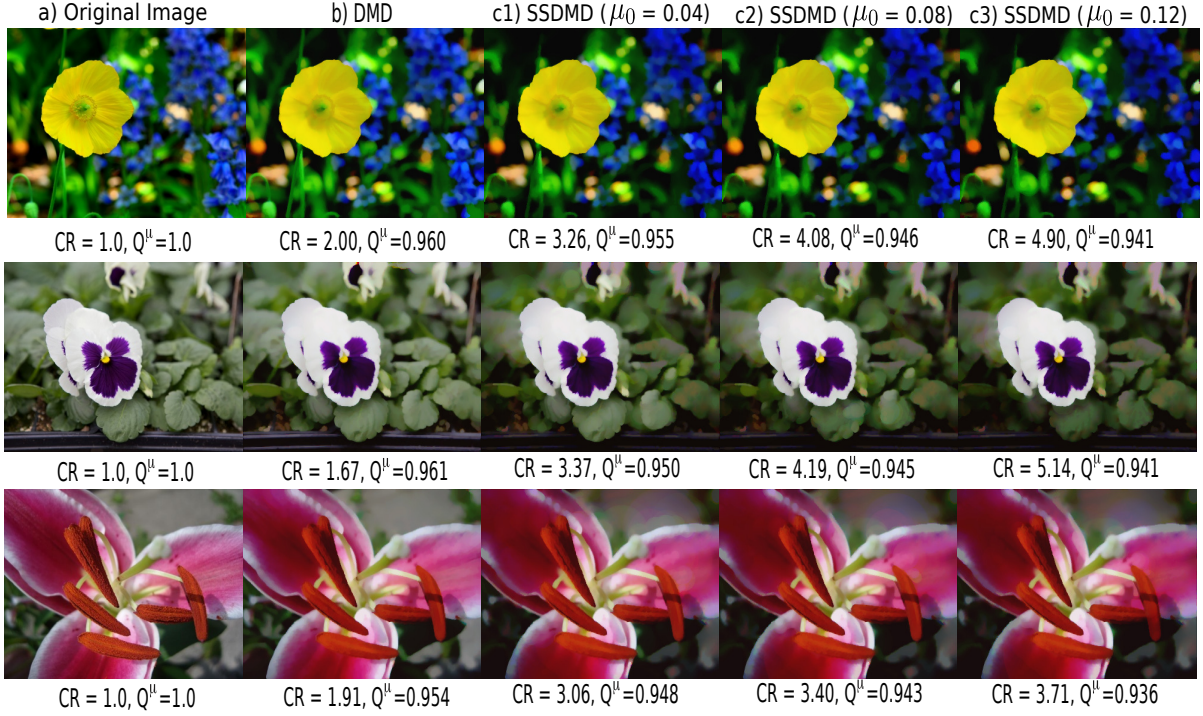


Figure 7: Progressive simplification control with the user-given threshold μ_0 .

three images. We see that the higher the quality setting, the higher the compression gain (green value) obtained by SSDMD as a preprocessor, except for the last row. In other words, for the same quality setting, SSDMD can *help* JPEG to increase compression rates for a minimal quality loss. This is explained by the fact that SSDMD removes small-scale sharp corners (which correspond to high frequencies in the image) in non-salient, background, image areas, thus making JPEG’s job overall easier.

5 DISCUSSION

We next discuss a few aspects of our proposed SSDMD method.

Genericity of the saliency map μ : In general, any saliency map that encodes which image areas are more important (salient) and which not for the application at hand can be used. In contrast to segmentation tasks, we do not require precise saliency maps. Figure 9(a-c) shows the SSDMD compression with three different saliency maps applied: DRFI (Jiang et al., 2013), SMD (Peng et al., 2017) and the very recent Iterative Saliency Estimator fLexible Framework (ITSELF) (de Melo Joao et al., 2020). ITSELF’s flexibility allows significant changes in the resulting



Figure 8: Comparison of JPEG (a) with the SSDMD method applied as preprocessor to JPEG (b) for three images. Green values show CR gains as compared to JPEG.

saliency maps by performing small adjustments to its parameters. Figure 9(c1) is a more nuanced version created by a relaxed threshold. Just like the result obtained by the DSR saliency map in Fig. 6,

all these three results get a higher compression ratio compared to the DMD method in Fig. 6 while maintaining similar quality. Besides, users can even customize the saliency maps themselves if the available saliency map μ does not meet their preferences. For this example, all these saliency maps say the stop traffic sign in the image is very important. Yet, if the user does not care about the sign, but rather wants to focus on the human face in the foreground, s/he could manually tune this area to be less than μ_0 , as shown in Fig. 9(d1), which is a user customization based on the DSR saliency map. This way, one can obtain a higher CR on the premise of meeting one’s quality requirements, as shown in Fig. 9(d2). We should stress again that what is a *good* saliency map is entirely at the user’s discretion and not a concern of SSDMD: Given a saliency map one is happy with, SSDMD compresses in low-saliency regions and preserves detail in high saliency regions.



Figure 9: SSDMD performance with different spatial saliency maps applied.

Ease of use: SSDMD can be used on any image, and adds two simple-to-control parameters: the

saliency-aware island threshold ϵ_0 and the spatially-regularized skeleton threshold μ_0 . These parameters have an intuitive meaning: ϵ_0 determines the scale of details that are kept in the image (higher values remove larger details); μ_0 controls how much the background/unimportant areas are simplified (higher values simplify background more).

Scalability: We inherit the speed of DMD (processing images up to 1000^2 pixels in a few milliseconds) given by the GPU-based MAT computation. Applying the saliency map involves only two simple additional thresholding operations.

Replicability: We provide the full source code of SSDMD, implemented in C++ and NVidia CUDA for replication purposes (Wang et al., 2020a).

Limitations: SSDMD cannot yet produce higher quality *and* better compression ratios than JPEG. Yet, as shown in Sec. 4, combining it with JPEG generically increases the latter’s compression while maintaining quality. Separately, the focus-and-context compression is only as good as the quality of the used saliency maps. When such maps incorrectly mark focus details as context, these will be simplified away; conversely, when context is marked as focus, the compression ratio will be suboptimal.

6 CONCLUSIONS

We have presented SSDMD, a method for saliency-aware image simplification and compression. SSDMD uses dense medial skeletons and a saliency map specifying which image areas can be simplified without compromising overall image perception. Additionally, we have proposed a saliency-dependent version of the MS-SSIM metric to evaluate SSDMD on images having a focus-and-context structure. Our results show that compared with the DMD method, SSDMD increases compression while keeping image quality high. SSDMD can also be used to improve the compression of standard JPEG though yield slightly lower quality. Currently, SSDMD is far from competing, standalone, with JPEG2000, HEVCIntra (Nguyen and Marpe, 2012), not to mention recent image compression methods that use deep learning (Toderici et al., 2016). However, this was not the goal of our paper. Rather, our purpose was to explore the *potential* of skeletons as an alternative tool to image representation. We believe that our results show that skeletons, when combined with saliency

maps, offer a promising tool for lossy image encoding, which can be refined next in the direction of competitive image compression.

We next aim to study more effective ways to encode skeletons prior to compression using piecewise-spline representations. Separately, we aim to test our method for simplifying general 2D and 3D scalar fields in scientific visualization, weighted with uncertainty-based saliency maps. In the long run, as outlined above, we believe that skeletons and saliency maps can provide effective and efficient tools for general-purpose, but also application-specific, lossy image representation.

Acknowledgements The first author acknowledges the China Scholarship Council (Grant number: 201806320354) for financial support.

REFERENCES

- Alaei, A., Raveaux, R., and Conte, D. (2017). Image quality assessment based on regions of interest. *Signal, Image and Video Processing*, 11:673–680.
- Andrushia, A. and Thangarajan, R. (2018). Saliency-based image compression using Walsh-Hadamard transform (WHT). In *Lecture Notes in Computational Vision and Biomechanics*, pages 21–42. Springer.
- Borji, A., Cheng, M., Jiang, H., and Li, J. (2015). Salient object detection: A benchmark. *IEEE TIP*, 24(12):5706–5722.
- Cazzato, D., Leo, M., Distanto, C., and Voos, H. (2020). When i look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking. *Sensors*, 20:3739.
- Chen, T., Cheng, M., Tan, P., Shamir, A., and Hu, S. (2009). Sketch2photo: Internet image montage. *ACM Trans. Graph.*, 28(5):1–10.
- Cheng, M. (2014). MSRA10K salient object database. <http://mmcheng.net/msra10k>.
- Cheng, M., Mitra, N. J., Huang, X., Torr, P. H., and Hu, S. (2014). Global contrast based salient region detection. *IEEE TPAMI*, 37(3):569–582.
- Costa, L. d. F. D. and Cesar, R. M. (2000). *Shape Analysis and Classification: Theory and Practice*. CRC Press, Inc., USA, 1st edition.
- de Melo Joao, L., de Castro Belem, F., and Falcao, A. X. (2020). Itself: Iterative saliency estimation flexible framework. Available at <https://arxiv.org/abs/2006.16956>.
- Engelke, U. and Le Callet, P. (2015). Perceived interest and overt visual attention in natural images. *Signal Processing: Image Communication*, 39:386–404.
- Falcão, A., Stolfi, J., and Lotufo, R. (2004). The image foresting transform: Theory, algorithms, and applications. *IEEE TPAMI*, 26(1):19–29.
- Goferman, S., Zelnik, L., and Tal, A. (2011). Context-aware saliency detection. *IEEE TPAMI*, 34(10):1915–1926.
- Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., and Li, S. (2013). Salient object detection: A discriminative regional feature integration approach. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2083–2090.
- Jiang, P., Ling, H., Yu, J., and Peng, J. (2013). Salient region detection by ufo: Uniqueness, focusness and objectness. In *Proc. ICCV*, pages 1976–1983.
- Koehoorn, J., Sobiecki, A., Boda, D., Diaconeasa, A., Doshi, S., Paisey, S., Jalba, A., and Telea, A. (2015). Automated digital hair removal by threshold decomposition and morphological analysis. In *Proc. ISMM*, volume 9082, pages 15–26. Springer.
- Koehoorn, J., Sobiecki, A., Rauber, P., Jalba, A., and Telea, A. (2016). Efficient and effective automated digital hair removal from dermoscopy images. *Math. Morphol. Theory Appl.*, 1.
- Le Callet, P. and Niebur, E. (2013). Visual attention and applications in multimedia technologies. *Proceedings of the IEEE*, 101(9):2058–2067.
- Li, C. and Bovik, A. (2010). Content-weighted video quality assessment using a three-component image model. *J. Electronic Imaging*, 19:110–130.
- Li, X., Lu, H., Zhang, L., Ruan, X., and Yang, M. (2013). Saliency detection via dense and sparse reconstruction. In *Proc. IEEE ICCV*, pages 2976–2983.
- Li, X., Lu, H., Zhang, L., Ruan, X., and Yang, M. (2013). Saliency detection via dense and sparse reconstruction. In *2013 IEEE International Conference on Computer Vision*, pages 2976–2983.
- Liu, H., Engelke, U., Wang, J., Callet, P., and Heynderickx, I. (2013). How does image content affect the added value of visual attention in objective image quality assessment? *IEEE Signal Processing Letters*, 20:355–358.
- Liu, H. and Heynderickx, I. (2011). Visual attention in objective image quality assessment: Based on eye-tracking data. *IEEE TCSVT*, 21(7):971–982.
- Liu, H. and Heynderickx, I. (2011). Visual attention in objective image quality assessment: Based on eye-tracking data. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(7):971–982.
- Movahedi, V. and Elder, J. H. (2010). Design and perceptual validation of performance measures for salient object segmentation. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 49–56.
- Nguyen, T. and Marpe, D. (2012). Performance analysis of HEVC-based intra coding for still image compression. In *Proc. IEEE Picture Coding Symp.*, pages 233–236.
- Nobuhara, H. and Hirota, K. (2004). Color image compression/reconstruction by yuv fuzzy wavelets. In *IEEE Annual Meeting of the Fuzzy Information, 2004. Processing NAFIPS '04.*, volume 2, pages 774–779.
- Peng, H., Li, B., Ling, H., Hu, W., Xiong, W., and Maybank, S. J. (2016). Salient object detection via structured matrix decomposition. *IEEE TPAMI*, 39(4):818–832.

- Peng, H., Li, B., Ling, H., Hu, W., Xiong, W., and Maybank, S. J. (2017). Salient object detection via structured matrix decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):818–832.
- Podpora, M. (2009). Yuv vs. rgb—a comparison of lossy compressions for human-oriented man-machine interfaces. *Zeszyty Naukowe. Elektryka*, pages 55–56.
- Podpora, M., Korbaś, G., and Kawala-Janik, A. (2014). Yuv vs rgb – choosing a color space for human-machine interaction. *Annals of Computer Science and Information Systems*, Vol. 3:29–34.
- Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., and Battisti, F. (2009). Tid2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10:30–45.
- Sheikh, H. R., Sabir, M. F., and Bovik, A. C. (2006). A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451.
- Shi, J., Yan, Q., Xu, L., and Jia, J. (2016). Hierarchical image saliency detection on extended cssd. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):717–729.
- Sobiecki, A., Koehoorn, J., Boda, D., Solovan, C., Diaconeasa, A., Jalba, A., and Telea, A. (2015). A new efficient method for digital hair removal by dense threshold analysis. In *Proc. 4th WC of Dermoscopy*. poster and extended abstract; Conference date: 21-04-2015.
- Telea, A. (2012). Feature preserving smoothing of shapes using saliency skeletons. In *Proc. VMLS*, pages 153–170.
- Telea, A. and van Wijk, J. (2002). An augmented fast marching method for computing skeletons and centerlines. In *Proc. VisSym*, pages 251–ff. Eurographics.
- Toderici, G., O’Malley, S., Hwang, S. J., Vincent, D., Minnen, D., Baluja, S., Covell, M., and Sukthankar, R. (2016). Variable rate image compression with recurrent neural networks. In *Proc. ICLR*. San Juan, Puerto Rico, May 2-4, 2016.
- Wang, J., de Melo João, L., Falcao, A., Kosinka, J., and Telea, A. (2020a). Implementation of SSDMD. <https://wangjieying.github.io/SSDMD-resources>.
- Wang, J., Terpstra, M., Kosinka, J., and Telea, A. (2020b). Quantitative evaluation of dense skeletons for image compression. *Information*, 11(5):274.
- Wang, Z. and Bovik, A. (2009). Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Proc Mag*, 26:98–117.
- Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13:600–612.
- Wang, Z., Simoncelli, E., and Bovik, A. (2003). Multiscale structural similarity for image quality assessment. In *Proc. Asilomar Conf. on Signals, Systems Computers*, pages 1398–1402.
- Zhang, J., Fang, S., Ehinger, K. A., Wei, H., Yang, W., Zhang, K., and Yang, J. (2018). Hypergraph optimization for salient region detection based on foreground and background queries. *IEEE Access*, 6:26729–26741.
- Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2011). Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386.
- Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2012). A comprehensive evaluation of full reference image quality assessment algorithms. In *2012 19th IEEE International Conference on Image Processing*, pages 1477–1480.
- Zünd, F., Pritch, Y., Sorkine-Hornung, A., Mangold, S., and Gross, T. (2013). Content-aware compression using saliency-driven image retargeting. In *2013 IEEE International Conference on Image Processing*, pages 1845–1849.
- Zwan, M. V. D., Meiburg, Y., and Telea, A. (2013). A dense medial descriptor for image analysis. In *Proc. VIS-APP*, pages 285–293.