Attention-guided Image Compression by Deep Reconstruction of Compressive Sensed Saliency Skeleton

Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

Xiaolin Wu[†]
McMaster Univeristy
xwu@ece.mcmaster.ca

Abstract

We propose a deep learning system for attention-guided dual-layer image compression (AGDL). In the AGDL compression system, an image is encoded into two layers, a base layer and an attention-guided refinement layer. Unlike the existing ROI image compression methods that spend an extra bit budget equally on all pixels in ROI, AGDL employs a CNN module to predict those pixels on and near a saliency sketch within ROI that are critical to perceptual quality. Only the critical pixels are further sampled by compressive sensing (CS) to form a very compact refinement layer. Another novel CNN method is developed to jointly decode the two compression layers for a much refined reconstruction, while strictly satisfying the transmitted CS constraints on perceptually critical pixels. Extensive experiments demonstrate that the proposed AGDL system advances the state of the art in perception-aware image compression.

1. Introduction

After decades of intensive research and development, visual signal compression techniques are approaching the rate-distortion performance limits. Any further significant improvements of bandwidth economy in visual communications have to come from smart human vision driven representations. In this direction the methodology of region-of-interest (ROI) image compression emerged about twenty years ago [7, 31, 3]. ROI compression is to exploit a well-known property of human vision: a viewer's attention is not evenly distributed in all parts of an image. Instead, our attentions focus on one or few regions of greater interests than the rest of the image, which pertain to salient foreground object(s). Background regions are delegated to our peripheral vision and hence have much lesser acuity. Playing this tapering of visual acuity away from ROIs, a ROI image compression method allocates a much lower bit bud-

get to encode pixels outside of ROIs than those inside, and saves a significant number of bits without materially sacrificing visual quality of compressed images.

In this work, we sharpen the existing tool of ROI image compression and propose a deep learning system of attention-guided dual-layer image compression (AGDL). In AGDL image compression, an image is encoded into two layers, a base layer I_b and an attention-guided refinement layer I_r. The base layer I_b is a conventional compressed image of low bit rate (high compression), such as those produced by JPEG, JPEG 2000, WebP, BPG, etc. The clarity of the base layer image just suffices to match the reduced level of acuity of peripheral vision. It is up to the additional attention-guided refinement layer I_r to boost the perceptual quality of ROI(s).

In existing ROI image compression methods, an extra bit budget is allocated to ROI and it is shared equally by all pixels in ROI. But on a second reflection, we should be more discriminating than ROI and spend extra bits only on pixels that can contribute to perceptual quality after being refined. Instead of a contiguous region of interest, we introduce a much sparser representation called saliency sketch to highlight semantically significant structures within ROI. One step further, we define a so-called critical pixel set that is the intersection of the saliency sketch and the set of pixels that have large reconstruction errors. The critical pixel set specifies a skeletal sub-image that needs to be further sampled and refined. For the saliency-driven refinement task, we take a more proactive approach than the straightforward CNN removal of compression artifacts [26, 15, 49, 44, 16, 14, 56, 58, 52, 54, 13, 17, 9]. In the AGDL system design, the encoder takes and transmits K additional samples of the critical pixel set in the form of compressive sensing (CS). The CS sampling produces novel critical information for the refinement layer, while having a very compact encoding of the novel information thanks to the small size of the critical pixel set.

The proposed AGDL image compression system needs to solve two key technical problems: 1. Detecting the saliency sketch and the critical pixels; 2. Refining the base

---

† Corresponding author.

layer with the CS measurements of the critical pixel set. The main technical contributions of this paper, besides the AGDL methodology, are the CNN solutions to the above two problems, one recognition and the other restoration.

2. Related Works

2.1. End-to-end optimized image compression

Toderici *et al*. [39] exploited recurrent neural networks for learned image compression. Some works [4, 38, 1] are proposed to approximate the non-differential quantization by a differentiable process to make the network end-to-end trainable. Toderici *et al*. [40] used recurrent neural networks (RNNs) to compress the residual information recursively. Rippel *et al*. [34, 2] proposed to learn the distribution of images using adversarial training to achieve better perceptual quality at extremely low bit rate. Li *et al*. [20] developed a method to allocate the content-aware bit rate under the guidance of a content-weighted importance map. [29, 5, 30, 19, 53] focused on investigating the adaptive context model for entropy estimation to achieve a better trade-off between reconstruction errors and required bits (entropy), among which the CNN methods of [30, 19] are the first to outperform BPG in PSNR.

2.2. ROI image compression

In the AGDL image compression system outlined above, the first step is to understand the image semantic composition and segment salient foreground objects. Detecting salient objects is a research topic in its own right. Recently, good progress has been made on this topic thanks to advances of deep learning research in computer vision, with a number of CNN segmentation methods published to extract salient objects from the background [24, 51, 42, 50, 10, 21, 33, 23, 46, 12, 43, 57, 27, 47, 11, 48, 45, 32]; they can be applied to compute ROIs. But for the purpose of image compression, we need to push further and seek for the shortest description of salient objects.

ROI based image compression, which is less discriminative than AGDL in selecting critical pixels for refinement, was an active research topic at the time of JPEG 2000 standard development [7, 31, 3]. Unlike JPEG that uses block DCT of very low spatial resolution (8×8 superpixel), JPEG 2000 is a two-dimensional wavelet representation and it can operate on images in good spatial resolution. This property makes ROI image compression possible. In conventional ROI coding, extra bits are spent to encode the ROI segment. As the ROI shape is determined by the contours of foreground objects, a flexible spatial descriptor inevitably consumes a significant amount of extra bandwidth. This cost of side information on ROI geometry could offset any rate-distortion performance gain made by ROI compression. This dilemma can be overcome by deep learn-

ing, as we demonstrate in the subsequent development of AGDL system and methods. By training a CNN to satisfactorily predict the saliency skeleton within ROI, AGDL compression strategy can enjoy the benefits of attention-guided compression free of side information.

Very recently, a CNN based ROI image compression method was published [6]. This is a pure CNN compression system of the standard auto-encoder architecture. The authors proposed the idea of extracting some CNN features specifically for the ROI. As explained in the introduction, the saliency sketch of AGDL is far more discriminative than a contiguous ROI; therefore it leads to more efficient use of extra refinement bits. Furthermore, we use CS measurements of critical pixels to exert input-specific constraints on the solution space of the underlying inverse problem, rather than solely relying on the statistics of the training set as in [6]. Finally, there is a drastic difference in encoder throughput between the method of [6] and our method. The base layer encoder of the proposed AGDL system can be any conventional image compressor (e.g. JPEG, JPEG 2000, WebP, BPG, etc.), which has a complexity orders of magnitude lower than CNN auto-encoder.

3. AGDL Compression System

In this section, we will introduce the design of the proposed AGDL image compression system, and two key technical contributions: 1. detecting saliency sketch and critical pixel set from the compressed base layer image; 2. Refining the base layer image with the CS measurements of the critical pixel set.

3.1. Overview

The overall framework of the proposed AGDL image compression system is shown in Fig. 1. It consists of a two-stage encoder and a joint decoder. Given an image I to be compressed, AGDL compression system first encodes I to a base layer I_b using a traditional image compressor, and then predicts the critical pixel mask $\mathbb{C}$ from the base layer I_b using a deep neural network $\mathcal{F}$. The resulting critical pixel mask $\mathbb{C}$ is used to extract the set of critical pixels $\mathbf{c}$. After that, AGDL system performs compressive sensing (CS) on the detected critical pixel set and transmits the CS measurements $\mathbf{y}$ along with the base layer I_b. The decoder takes the base layer I_b and the CS measurements $\mathbf{y}$ of the critical pixel set as input to produce a refined image $\hat{I}$ with highlighted semantic structures by a restoration network $\mathcal{G}$ and a CS refining module $\mathcal{R}$.

3.2. Saliency sketch and critical pixels

Existing ROI image compression methods, including the recently proposed pure CNN ROI compression system [6], weigh all pixels in ROI equally. However, not all pixels in
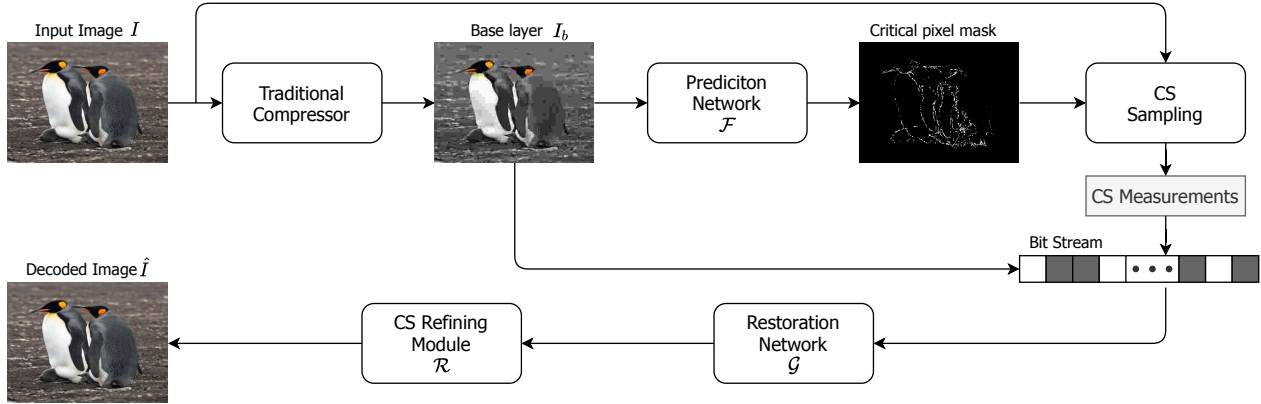
Figure 1: The overall framework of the proposed AGDL image compression system.

ROI carry the same significance to visual quality. For example in Fig 2b, the featureless power portions of the three baskets matter much less to visual perception than the textured upper portions. A rate-distortion more efficient way of coding is to allocate more bits only to pixel structures that contribute the most to improving perceptual quality, such as edges and textures. To this end, we introduce a much sparser presentation than ROI, called saliency sketch, which is defined as the edge map of the object(s) in ROI, as shown in Fig. 2c.

In fact, we can be even more selective than saliency sketch, if considering the recent progresses made on CNN based compression artifact removal (CAR) techniques [49, 44, 56, 9]. These learning methods can restore many pixels belonging to saliency sketch, and the CNN recoverable pixels need not be additionally sampled and transmitted. Thus the AGDL encoder only needs to send new information on the pixels that belong to saliency sketch and but also have large reconstruction error. We define these pixels critical pixels.

Denoting the edge skeleton of I by Ω_s, the ROI of I by Ω_i, the set of pixels of large reconstruction errors after CAR by Ω_e, then the critical pixel mask $\mathbb{C}$ can be represented as:

$$\mathbb{C} = \Omega_s \cap \Omega_i \cap \Omega_e \tag{1}$$

In Fig. 2d, the critical pixel mask indicates the locations of the critical pixels. The critical pixel set specifies a skeletal sub-image that needs to be further sampled and refined.

3.3. Detecting critical pixel set

In traditional ROI image coding, the ROI geometry is explicitly encoded and therefore is a part of compression code stream. The extra bits required to transmit the ROI shape could offset any rate-distortion performance gain made by ROI compression. This dilemma can be overcome by deep learning if a CNN can learn to predict the ROI mask from
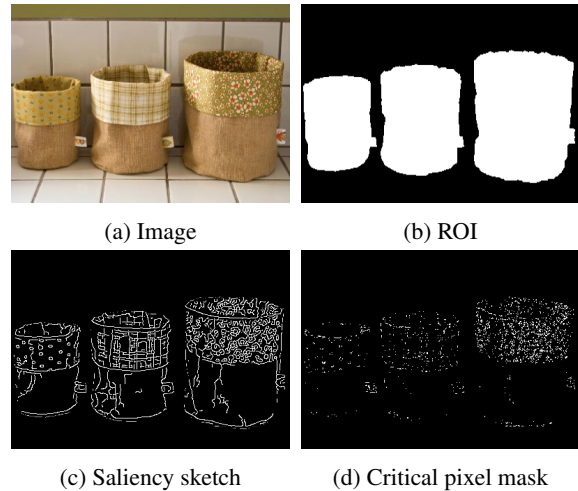

Figure 2: Examples of the natural image, ROI map, and the proposed saliency sketch and critical pixel mask.

the base layer image I_b. This eliminates the need to transmit the ROI shape because the decoder can make the same ROI prediction as the encoder.

In the AGDL image compression system, we push further and drive a CNN $\mathcal{F}$ to predict the critical pixel mask $\mathbb{C}$ that is a subset of ROI from the base layer I_b. This learning task is more demanding, but it is nevertheless feasible because the critical pixel mask $\mathbb{C}$ of an image can be computed to generate paired data for supervised learning. This is a strategy of squeezing out coding gains by computation power and big data.

Specifically, we adopt an existing CAR network called DnCNN [49] to initially restore base layer I_b and then identify the set Ω_e of those pixels that still have large restoration errors. In addition, we use a salient object network BASNet [33] to calculate the ROI region Ω_i, and detect the edge skeleton Ω_s using Canny operator.
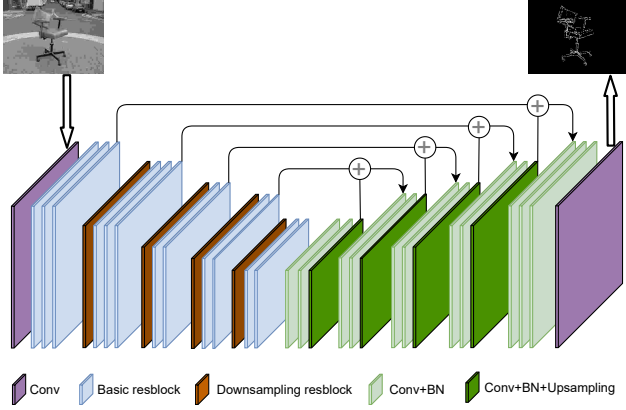
Figure 3: Architecture of the proposed critical pixel mask prediction network $\mathcal{F}$.

Given Ω_i, Ω_s and Ω_e, the critical pixel mask $\mathbb{C}$ is determined. So we can build paired data (baser layer images I_b and the corresponding critical pixel masks $\mathbb{C}$) to train the prediction network $\mathcal{F}$. Let $\mathcal{F}$ be the prediction network:

$$\mathbb{C} = \mathcal{F}(I_b) \tag{2}$$

The architecture of the proposed prediction network $\mathcal{F}$ is revised from BASNet [33], a network designed for salient object detection. As shown in Fig. 3, the prediction network $\mathcal{F}$ is a U-Net-like Encoder-Decoder network [35], which learns to predict critical pixel mask from base layer image. We design the critical pixel mask prediction network as an Encoder-Decoder architecture because it is able to capture high level global contexts and low level details at the same time [35, 28]. The encoder part has an input convolution layer and five stages comprised of residual blocks. The input layer has 64 convolution filters with size of 3×3 and stride of 1. The first stage is size-invariant and the other four stages gradually reduce the feature map resolution by downsampling resblocks to obtain a larger receptive field. The decoder is almost symmetrical to the encoder. Each stage consists of three convolution layers followed by a batch normalization and a ReLU activation function. The input of each layer is the concatenated feature maps of the upsampled output from its previous layer and its corresponding layer in the encoder.

The critical pixel set $\mathbf{c}$ can be extracted based on the predicted critical pixel mask $\mathbb{C}$, and then rearranged into a column vector. After that, AGDL compression system performs Compressed sensing on the critical pixel set $\mathbf{c}$ with a full row rank, fat CS sampling matrix H (far fewer rows than columns):

$$\mathbf{y} = H \cdot \mathbf{c} \tag{3}$$

where $\mathbf{y}$ is the CS measurements of the critical pixel set.

The CS measurements $\mathbf{y}$ and base layer I_b will be transmitted to the decoder end.

3.4. Dual-layer joint decoding

The most important and technically involved component of the AGDL image compression system is its CNN decoder. The task of AGDL decoding is to refine the JPEG-coded base layer I_b, aided by the CS-coded side information on saliency skeleton. Specifically, the AGDL decoder receives the base layer I_b and refinement layer I_r (CS measurements of critical pixels), and then jointly decodes the two layers to produce a refined image $\hat{I}$ which strictly satisfies the CS constraints. In essence, the AGDL decoder is a heavy-duty CNN that removes the compression artifacts of the base layer image I_b with encoder-supplied strong priors on ROI.

By satisfying the CS constraints we mean that after the critical pixel set $\hat{c}$ in the CNN refined image $\hat{I}$ is sampled by the CS sampling matrix H, the resulting CS measurements $\hat{y}$ equal to the received measurements $\mathbf{y}$, that is

$$H \cdot \hat{\mathbf{c}} = \mathbf{y} \tag{4}$$

To the best of our knowledge, we are the first to impose such constraints on CNN outputs. This way of confining the solution space of an inverse problem in CNNs poses a technical challenge. We overcome the difficulty by cascading a restoration network $\mathcal{G}$ and a CS refining module $\mathcal{R}$, in which the latter constrains the output of the former by the CS measurements. The joint decoding process can be formulate as:

$$I_g = \mathcal{G}(I_b) \tag{5}$$
$$\hat{I} = \mathcal{R}(I_g, \mathbf{y}) \tag{6}$$

where I_b is the decoded result of a traditional image compressor (decompressed image); the restoration network $\mathcal{G}$ performs a post-processing on I_b, called soft decoding, aiming to remove compression artifacts in I_b. The result of soft decoding is a restored image I_g. The final step of the AGDL system is to adjust the set of critical pixels in I_g, denoted by $\mathbf{c}_g$, so that their values strictly satisfy the set of K CS measurements. Among all possible such K-dimensional adjustment vectors δ, the one δ_* of the minimum ℓ_2 norm generates the final refinement image $\hat{I} = I_g + \delta_*$.

Next we develop the CS refining module $\mathcal{R}$ that imposes constraints on the final output of the AGDL system. Firstly, I_g must not satisfy the CS constraint, that is

$$H \cdot \mathbf{c}_g \neq \mathbf{y} \tag{7}$$

where $\mathbf{c}_g$ is the critical pixel set in the restored image I_g. We hope to make the minimum adjustment to the output image
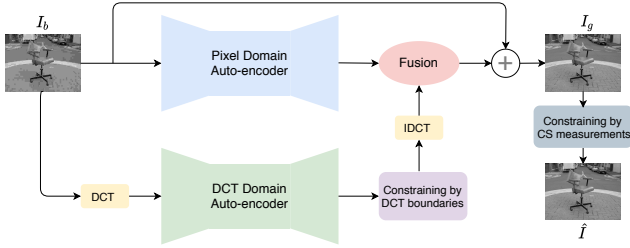
Figure 4: Architecture of the decoder, including the restoration network $\mathcal{G}$ and the CS refining module $\mathcal{R}$.

I_g (or to the critical pixel set $\mathbf{c}_g$) so that the adjusted image can satisfy the CS constraint. This forms the following optimization problem:

$$\text{minimize} \quad ||\delta|| \tag{8}$$
$$\text{subject to} \quad H \cdot (\mathbf{c}_g + \delta) = \mathbf{y} \tag{9}$$

Since the CS sampling matrix H is full row rank, so the above optimization problem has the solution that is:

$$\delta^* = H^T (HH^T)^{-1} \cdot (\mathbf{y} - H \cdot \mathbf{c}_g) \tag{10}$$

This is the classical least-norm solution of undetermined equations. Detailed solving steps will be given in the supplementary material. Let $\hat{\mathbf{c}} = \mathbf{c}_g + \delta^*$, so the adjusted critical pixel set $\hat{\mathbf{c}}$ satisfies the CS constraint. It is noteworthy that the adjustment is linear, so it can participate in the back propagation.

In the design of the restoration network $\mathcal{G}$, we adopt a dual-domain (pixel domain and transform domain) network to take full advantage of redundancies in both pixel and transform domains [56, 55]. In most traditional image compression methods, images are converted to a transform domain (e.g., DCT, wavelet, etc.) and then quantized. The encoder prior information contained in the transform domain can help improve the performance of soft-decoding.

The base layer of the AGDL system can be any of existing image compression methods. In this paper, we choose JPEG as the base layer to develop the restoration network $\mathcal{G}$, as it is the most common compression method. As shown in Fig. 4, the proposed restoration network $\mathcal{G}$ has two branches, one operating in pixel-domain and the other in DCT domain. The pixel-domain branch is to restore the pixel values directly, while the DCT-domain branch aims to recover the DCT coefficients of the ground truth. The fusion network combines these two branches to produce the restored image I_g. After CS refinement, $\hat{I}$ is used to calculate loss to optimize the network $\mathcal{G}$.

Now we are at the point to present the overall pipeline of AGDL compression system in Algorithm. 1.

Algorithm 1 Framework of AGDL compression system.

Input: The original image, I;
Output: The decoded image, $\hat{I}$;
Encoding:
1: Encoding I into a base layer I_b using JPEG;
2: Predicting critical pixel mask $\mathbb{C}$ from the base layer I_b by the prediction network $\mathcal{F}$, $\mathbb{C} = \mathcal{F}(I_b)$;
3: Extracting critical pixel set $\mathbf{c}$ based on $\mathbb{C}$;
4: Applying compressive sensing on $\mathbf{c}$, $\mathbf{y} = H \cdot \mathbf{c}$;
5: Transmitting I_b and $\mathbf{y}$;
Decoding:
1: Soft-decoding I_b by the network $\mathcal{G}$, $I_g = \mathcal{G}(I_b)$;
2: Calculating the minimum adjustment to satisfy the CS constraint, $\delta^* = H^T(HH^T)^{-1} \cdot (\mathbf{y} - H \cdot \mathbf{c}_g)$;
3: Applying the adjustment, $\hat{I} = I_g + \delta^*$;
4: Output the final refinement image $\hat{I}$;

---

4. Experiments

In this section, we introduce the implementation details of the proposed AGDL image compression system. To systematically evaluate and analyze the performance of the AGDL compression system, we conduct extensive experiments on two scenarios: portrait and general objects, and compare our results with several stat-of-the-art methods.

4.1. Dataset

Portrait. We adopt the portrait dataset provided by Shen et al. [37] for training and evaluation. It contains 2000 images of 600×800 resolution where 1700 and 300 images are split as training and testing set respectively. To overcome the lack of training data, we augment images by utilizing rotation and left-right flip, as suggested in [36]. Each training image is rotated by $[-15°, 15°]$ in steps of $5°$ and left-right flipped, which means that a total of 23800 training images are obtained.

General objects. In the scenario for general objects, we adopt the DUTS dataset [41] for training and testing. Currently, DUTS is the largest and most frequently used dataset for salient object detection. DUTS dataset consists of two parts: DUTS-TR and DUTS-TE. DUTS-TR contains 10553 images in total. We augment this dataset by horizontal flipping to obtain 21106 training images. DUTS-TE, which contains 5019 images, is selected as our evaluation dataset.

All these images are resized to 300×400 resolution for training and evaluation. We choose JPEG as the traditional image compressor of the AGDL system, as JPEG is the most widely used image compression method. For both scenarios, we compress the images using JPEG with quality factor in $[10, 100]$ in steps of 10 to form a multi-rate training set. All training and evaluation processes are performed on the luminance channel (in YCbCr color space).
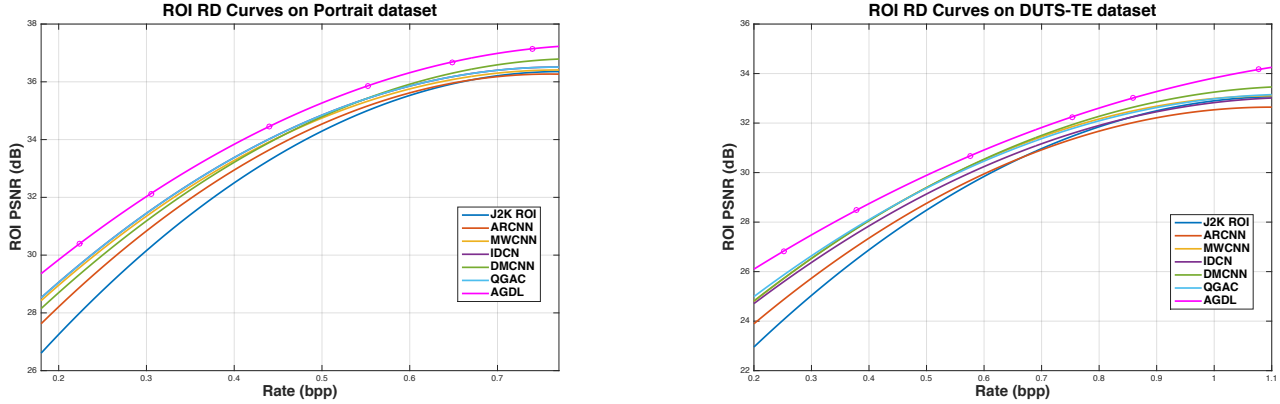
Figure 5: ROI RD curves of the competing methods on Portrait and DUTS-TE datasets.

4.2. Training details

Totally, we have two networks to train, a prediction network $\mathcal{F}$ and a restoration network $\mathcal{G}$. Next, we introduce the training details of the two networks separately.

Prediction network $\mathcal{F}$. To train the network $\mathcal{F}$ for predicting critical pixel mask, we first adopt DnCNN [49] to initially restore the JPEG-coded images and then identify the set Ω_e of those pixels that still have large restoration errors (error > 8). In addition, we use a salient object network BASNet [33] to calculate the ROI region Ω_i, and detect the edge skeleton Ω_s using Canny operator. Then, we get the critical pixel mask $\mathbb{C}$ according to Eq. 1. The critical pixel mask $\mathbb{C}$ is a binary mask, in which 1 means the current location is critical pixel and 0 vice versa. The prediction network $\mathcal{F}$ takes JPEG-coded images as input and outputs the corresponding critical pixel masks, so it solves a binary classification problem for each pixel location. To this end, we train $\mathcal{F}$ using the Binary Cross Entropy (BCE) loss function. When inferring, prediction network $\mathcal{F}$ outputs a probability value in $[0, 1]$ for each pixel location, indicating the probability of being a critical pixel. Top K pixels ranked by probability form the critical pixel set to be further sampled and transmitted. More details about the CS sampling matrix H are given in the supplementary material.

Restoration network $\mathcal{G}$. To reduce the risk of overfitting, the restoration network $\mathcal{G}$ is pretrained using the DIV2K [22] and Flikr2K [22] datasets. After pretrained, the restoration network $\mathcal{G}$ is fine-tuned on portrait dataset [37] and DUTS-TR [41] separately, under the constraints of the CS measurements. $\mathcal{L}_1$ loss is adopted to optimize the restoration network $\mathcal{G}$.

All training processes use the Adam [18] optimizer by setting $\beta_1 = 0.9$ and $\beta_2 = 0.999$, with a batch size of 16. The network is trained with 100 epochs at the learning rate of 10^{-4} and other epochs with learning rate of 10^{-5}. The algorithms are implemented in the MindSpore framework.

4.3. Comparison with state-of-the-art methods

To demonstrate the advantages of the proposed AGDL compression system, we compare AGDL with several other compression systems, in which JPEG is also used as the compressor and several deep-learning based compression artifact reduction methods ARCNN [8], MWCNN [25], IDCN [58], DMCNN [56], QGAC [9] are used as the soft decoder. In order to factor out the effects of different training sets and conduct a fair comparison, we fine-tune all CNN networks in the comparison group using the same datasets (Portrait and DUTS) in our experiments. We also compare AGDL with JPEG2000's ROI coding which is implemented in Kakadu JPEG2000 software. In the AGDL system, the total bit rates need to be transmitted are the sum of the rates of JPEG-coded base layer and the CS-coded side information. To facilitate fair rate-distortion performance evaluations, for each test image, the rates of the competing compression systems are adjusted to match or be slightly higher than that of the AGDL compression system.

Quantitative results. We present rate distortion (RD) curves of ROI in Fig. 5. The rate is calculated by bits consumed to encode the entire image averaged per pixel (bpp), and the distortion is measured by the PSNR of the ROI area. For AGDL, the rate is the sum of the bits consumed by the JPEG-coded base layer and the CS-coded side information. As shown in Fig. 5, the proposed AGDL compression system outperforms all the competing methods by a large margin, on both portrait images and general object images. For portrait images, the PSNR gain obtained by AGDL is relatively uniform in bit rate. However, for general objects, the PSNR gain is unevenly distributed. Specifically, the more extreme the bit rate, the greater the PSNR gain.

Qualitative results. In addition to the quantitative results of RD curves, we also present the visual comparisons of different methods, as shown in Fig. 6 and 7. QGAC [9] is the state-of-the-art CNN method for compression artifacts
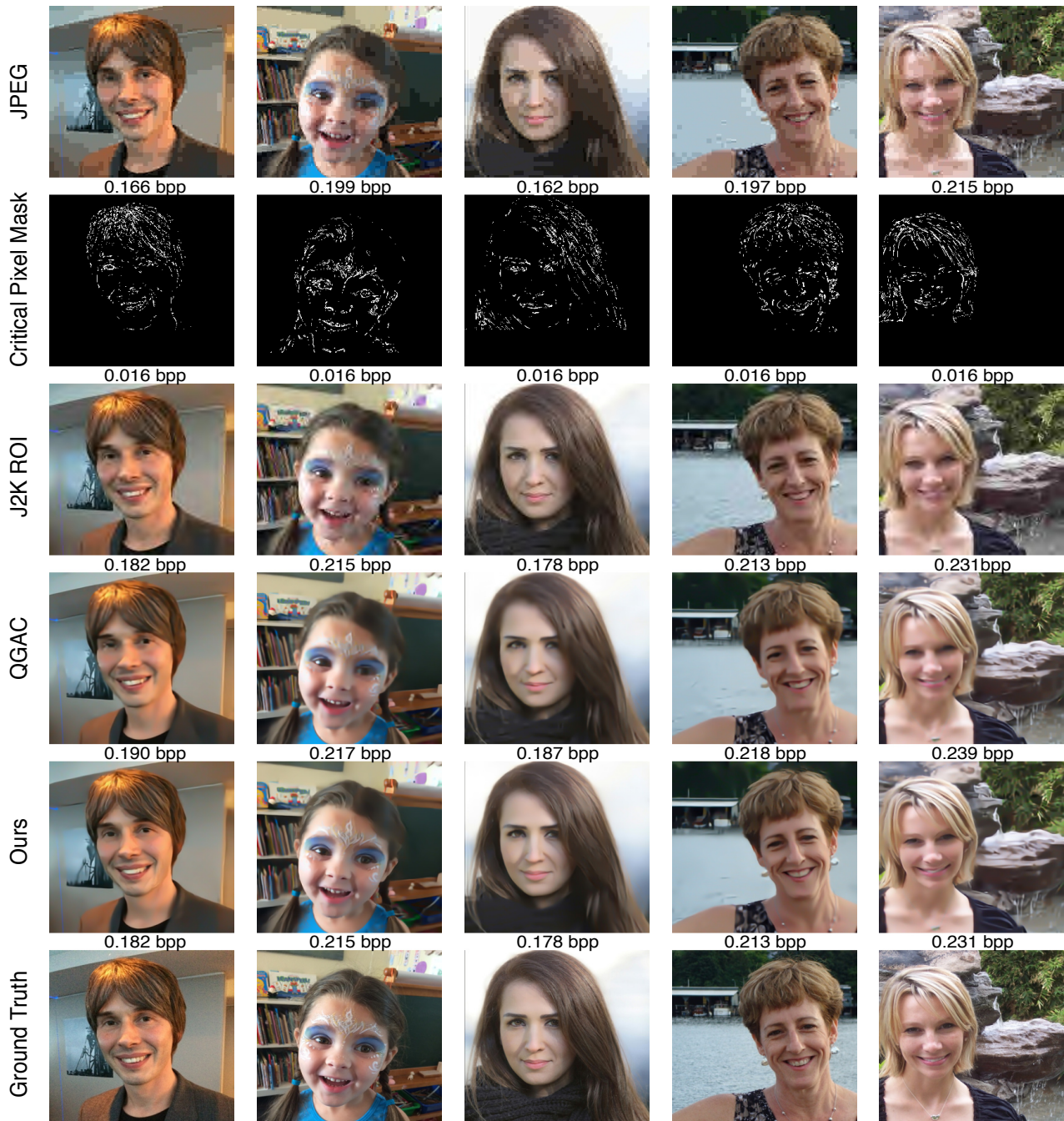
Figure 6: Visual comparisons of different methods on portraits.

reduction, so we only show QGAC's results for visual comparison due to page limit. The complete visual comparisons of all competing methods will be given in the supplementary material. In the visual comparisons, we add the color channels (CbCr) back for the best visual quality. In Fig. 6, we can see that the AGDL compression system can preserve facial features better than the state-of-the-art QGAC method and J2K ROI compression (note clearer eyes and hair, sharper muscle contours). For general objects, Fig. 7 shows us that the AGDL system is able to preserve the small structures with the help of CS constraints, such as the spots on the sika deer and the lines on the butterfly. In addition, AGDL can make animal hair more realistic, while QGAC makes the hair look too smooth.
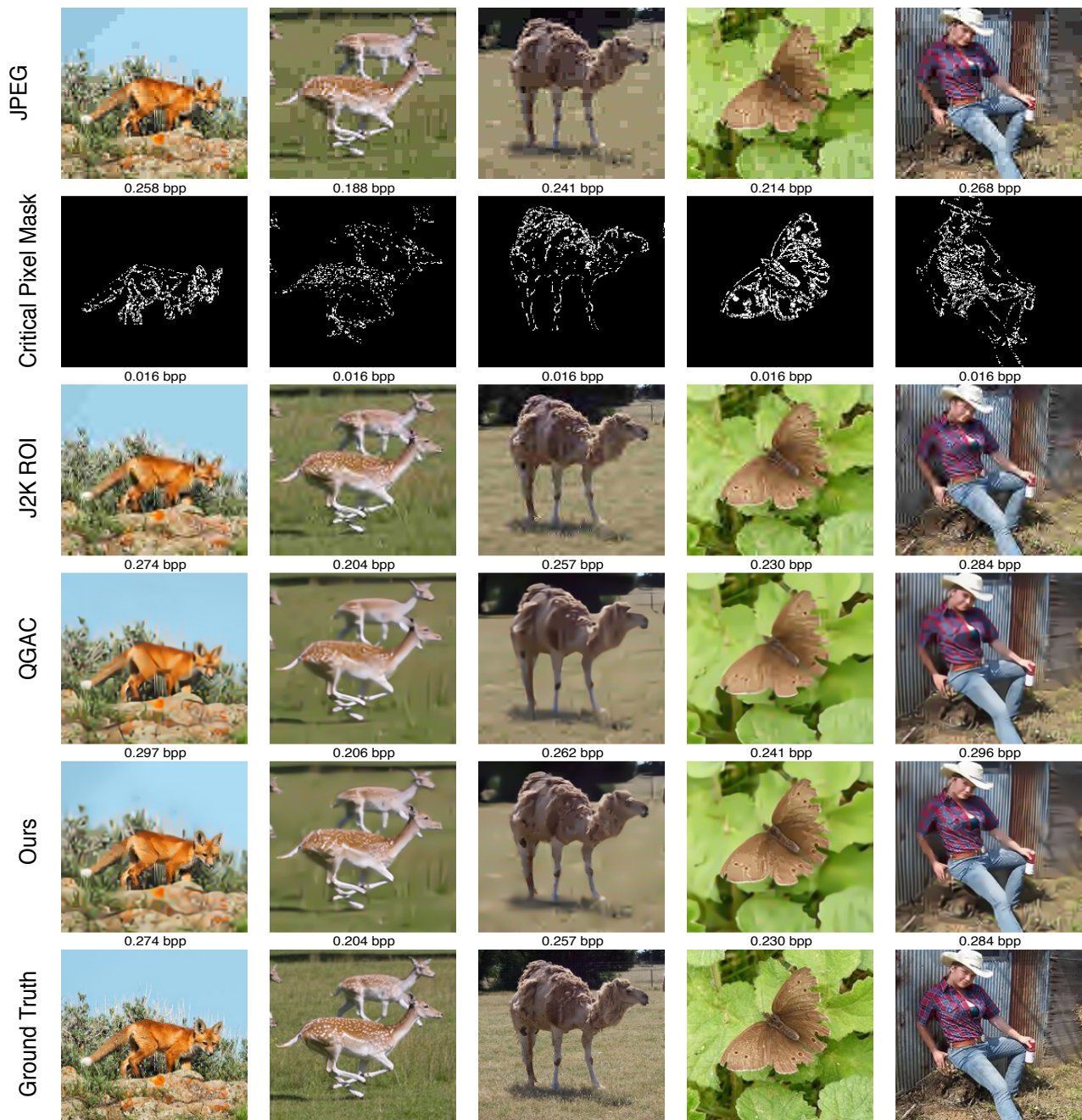
Figure 7: Visual comparisons of different methods on general objects.

5. Conclusion

We present a deep learning system AGDL for attention-guided dual-layer image compression. AGDL employs a CNN module to predict those pixels on and near a saliency sketch within ROI that are critical to perceptual quality. Only the critical pixels are further sampled by compressive sensing. In addition, AGDL jointly decodes the two compression code layers for a much refined reconstruction, while strictly satisfying the transmitted CS constraints on perceptually critical pixels.

Acknowledgments

References

[1] Eirikur Agustsson, Fabian Mentzer, Michael Tschannen, Lukas Cavigelli, Radu Timofte, Luca Benini, and Luc V Gool. Soft-to-hard vector quantization for end-to-end learning compressible representations. In *Advances in Neural Information Processing Systems*, pages 1141–1151, 2017. 2

[2] Eirikur Agustsson, Michael Tschannen, Fabian Mentzer, Radu Timofte, and Luc Van Gool. Generative adversarial networks for extreme learned image compression. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 221–231, 2019. 2

[3] Eiji Atsumi and Nariman Farvardin. Lossy/lossless region-of-interest image coding based on set partitioning in hierarchical trees. In *International Conference on Image Processing (ICIP)*, volume 1, pages 87–91. IEEE, 1998. 1, 2

[4] Johannes Ballé, Valero Laparra, and Eero P Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016. 2

[5] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436*, 2018. 2

[6] Chunlei Cai, Li Chen, Xiaoyun Zhang, and Zhiyong Gao. End-to-end optimized roi image compression. *IEEE Transactions on Image Processing*, 29:3442–3457, 2019. 2

[7] Charilaos Christopoulos, Joel Askelof, and Mathias Larsson. Efficient methods for encoding regions of interest in the upcoming jpeg2000 still image coding standard. *IEEE Signal Processing Letters*, 7(9):247–249, 2000. 1, 2

[8] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang. Compression artifacts reduction by a deep convolutional network. In *IEEE International Conference on Computer Vision (ICCV)*, pages 576–584, 2015. 6

[9] Max Ehrlich, Ser-Nam Lim, Larry Davis, and Abhinav Shrivastava. Quantization guided jpeg artifact correction. In *European Conference on Computer Vision (ECCV*. Springer, 2020. 1, 3, 6

[10] Deng-Ping Fan, Ming-Ming Cheng, Jiang-Jiang Liu, Shang-Hua Gao, Qibin Hou, and Ali Borji. Salient objects in clutter: Bringing salient object detection to the foreground. In *European conference on computer vision (ECCV)*, pages 186–202, 2018. 2

[11] Deng-Ping Fan, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Huazhu Fu, and Ming-Ming Cheng. Taking a deeper look at co-salient object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2

[12] Mengyang Feng, Huchuan Lu, and Errui Ding. Attentive feedback network for boundary-aware salient object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1623–1632, 2019. 2

[13] Xueyang Fu, Zheng-Jun Zha, Feng Wu, Xinghao Ding, and John Paisley. Jpeg artifacts reduction via deep convolutional sparse coding. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2501–2510, 2019. 1

[14] Leonardo Galteri, Lorenzo Seidenari, Marco Bertini, and Alberto Del Bimbo. Deep generative adversarial compression artifact removal. *arXiv preprint arXiv:1704.02518*, 2017. 1

[15] Jun Guo and Hongyang Chao. Building dual-domain representations for compression artifacts reduction. In *European Conference on Computer Vision (ECCV*, pages 628–644. Springer, 2016. 1

[16] Jun Guo and Hongyang Chao. One-to-many network for visually pleasing compression artifacts reduction. *arXiv preprint arXiv:1611.04994*, 2016. 1

[17] Zhi Jin, Muhammad Zafar Iqbal, Wenbin Zou, Xia Li, and Eckehard Steinbach. Dual-stream multi-path recursive residual network for jpeg image compression artifacts reduction. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020. 1

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[19] Jooyoung Lee, Seunghyun Cho, and Seung-Kwon Beack. Context-adaptive entropy model for end-to-end optimized image compression. *arXiv preprint arXiv:1809.10452*, 2018. 2

[20] Mu Li, Wangmeng Zuo, Shuhang Gu, Debin Zhao, and David Zhang. Learning convolutional networks for content-weighted image compression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3214–3223, 2018. 2

[21] Xin Li, Fan Yang, Hong Cheng, Wei Liu, and Dinggang Shen. Contour knowledge transfer for salient object detection. In *European Conference on Computer Vision (ECCV)*, pages 355–370, 2018. 2

[22] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 136–144, 2017. 6

[23] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang. A simple pooling-based design for real-time salient object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3917–3926, 2019. 2

[24] Nian Liu, Junwei Han, and Ming-Hsuan Yang. Picanet: Learning pixel-wise contextual attention for saliency detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3089–3098, 2018. 2

[25] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. 6

[26] Xianming Liu, Xiaolin Wu, Jiantao Zhou, and Debin Zhao. Data-driven sparsity-based restoration of jpeg-compressed images in dual transform-pixel domain. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5171–5178, 2015. 1

[27] Yi Liu, Qiang Zhang, Dingwen Zhang, and Jungong Han. Employing deep part-object relationships for salient object detection. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 2

[28] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015. 4

[29] Fabian Mentzer, Eirikur Agustsson, Michael Tschannen, Radu Timofte, and Luc Van Gool. Conditional probability models for deep image compression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4394–4402, 2018. 2

[30] David Minnen, Johannes Ballé, and George D Toderici. Joint autoregressive and hierarchical priors for learned image compression. In *Advances in Neural Information Processing Systems*, pages 10771–10780, 2018. 2

[31] David Nister and Charilaos Christopoulos. Lossless region of interest with a naturally progressive still image coding algorithm. In *International Conference on Image Processing (ICIP)*, pages 856–860. IEEE, 1998. 1, 2

[32] Youwei Pang, Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu. Multi-scale interactive network for salient object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2

[33] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7479–7489, 2019. 2, 3, 4, 6

[34] Oren Rippel and Lubomir Bourdev. Real-time adaptive image compression. *arXiv preprint arXiv:1705.05823*, 2017. 2

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4

[36] Seokjun Seo, Seungwoo Choi, Martin Kersner, Beomjun Shin, Hyungsuk Yoon, Hyeongmin Byun, and Sungjoo Ha. Towards real-time automatic portrait matting on mobile devices. *arXiv preprint arXiv:1904.03816*, 2019. 5

[37] Xiaoyong Shen, Xin Tao, Hongyun Gao, Chao Zhou, and Jiaya Jia. Deep automatic portrait matting. In *European conference on computer vision (ECCV)*, pages 92–107. Springer, 2016. 5, 6

[38] Lucas Theis, Wenzhe Shi, Andrew Cunningham, and Ferenc Huszár. Lossy image compression with compressive autoencoders. *arXiv preprint arXiv:1703.00395*, 2017. 2

[39] George Toderici, Sean M O'Malley, Sung Jin Hwang, Damien Vincent, David Minnen, Shumeet Baluja, Michele Covell, and Rahul Sukthankar. Variable rate image compression with recurrent neural networks. *arXiv preprint arXiv:1511.06085*, 2015. 2

[40] George Toderici, Damien Vincent, Nick Johnston, Sung Jin Hwang, David Minnen, Joel Shor, and Michele Covell. Full resolution image compression with recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5306–5314, 2017. 2

[41] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *CVPR*, 2017. 5, 6

[42] Tiantian Wang, Lihe Zhang, Shuo Wang, Huchuan Lu, Gang Yang, Xiang Ruan, and Ali Borji. Detect globally, refine locally: A novel approach to saliency detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3127–3135, 2018. 2

[43] Wenguan Wang, Shuyang Zhao, Jianbing Shen, Steven CH Hoi, and Ali Borji. Salient object detection with pyramid attention and salient edges. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR*, pages 1448–1457, 2019. 2

[44] Zhangyang Wang, Ding Liu, Shiyu Chang, Qing Ling, Yingzhen Yang, and Thomas S Huang. D3: Deep dual-domain based fast restoration of jpeg-compressed images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2764–2772, 2016. 1, 3

[45] Jun Wei, Shuhui Wang, Zhe Wu, Chi Su, Qingming Huang, and Qi Tian. Label decoupling framework for salient object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2

[46] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3907–3916, 2019. 2

[47] Yi Zeng, Pingping Zhang, Jianming Zhang, Zhe Lin, and Huchuan Lu. Towards high-resolution salient object detection. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 2

[48] Jing Zhang, Xin Yu, Aixuan Li, Peipei Song, Bowen Liu, and Yuchao Dai. Weakly-supervised salient object detection via scribble annotations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2

[49] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 1, 3, 6

[50] Lu Zhang, Ju Dai, Huchuan Lu, You He, and Gang Wang. A bi-directional message passing model for salient object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1741–1750, 2018. 2

[51] Xiaoning Zhang, Tiantian Wang, Jinqing Qi, Huchuan Lu, and Gang Wang. Progressive attention guided recurrent network for salient object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 714–722, 2018. 2

[52] Xi Zhang and Xiaolin Wu. Near-lossless ℓ_∞-constrained image decompression via deep neural network. In *2019 Data Compression Conference (DCC)*, pages 33–42. IEEE, 2019. 1

[53] Xi Zhang and Xiaolin Wu. Nonlinear prediction of multidimensional signals via deep regression with applications to image coding. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1602–1606. IEEE, 2019. 2

[54] Xi Zhang and Xiaolin Wu. Ultra high fidelity deep image decompression with ℓ_∞-constrained compression. *IEEE Transactions on Image Processing*, 30:963–975, 2020. 1

[55] Xi Zhang, Xiaolin Wu, Xinliang Zhai, Xianye Ben, and Chengjie Tu. Davd-net: Deep audio-aided video decompression of talking heads. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 5

[56] Xiaoshuai Zhang, Wenhan Yang, Yueyu Hu, and Jiaying Liu. Dmcnn: Dual-domain multi-scale convolutional neural network for compression artifacts removal. In *IEEE International Conference on Image Processing (ICIP)*, pages 390–394. IEEE, 2018. 1, 3, 5, 6

[57] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet: Edge guidance network for salient object detection. In *IEEE International Conference on Computer Vision (ICCV)*, pages 8779–8788, 2019. 2

[58] Bolun Zheng, Yaowu Chen, Xiang Tian, Fan Zhou, and Xuesong Liu. Implicit dual-domain convolutional network for robust color image compression artifact reduction. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019. 1, 6