IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON AI-DRIVEN BIG DATA PROCESSING: THEORY, METHODOLOGY, AND APPLICATIONS

Received November 1, 2018, accepted December 3, 2018, date of publication December 6, 2018,
date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885347

CarvingNet: Content-Guided Seam Carving Using Deep Convolution Neural Network

EUNGYEOL SONG, MINKYU LEE, AND SANGYOUN LEE, (Member, IEEE)
Department of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea

Corresponding author: Sangyoun Lee (syleee@yonsei.ac.kr)

This work was supported by the Institute for Information and Communications Technology Promotion (IITP) Grant funded by the Korean Government (MSIP) (Development of the High-Precision Natural 3D View Generation Technology Using Smart-Car Multi-Sensors and Deep Learning) under Grant 2016-0-00197.

ABSTRACT We propose an improved content-aware image resizing method that uses deep learning. The proposed method is extended from seam carving, which is another image resizing method. Seam carving uses the energy map from an image. It also removes a seam where the energy is the minimum. We propose a method for creating a deep energy map using an encoder–decoder convolution neural network. A deep energy map preserves important parts or boundaries in an image, without distortion. Furthermore, it has the characteristic that uniform intensity of edges is displayed for all images. Four well-known resizing methods and our proposed method were evaluated in terms of aspect ratio similarity. In such an objective evaluation, the proposed method demonstrated better results than the other four algorithms. Our proposed method can reduce the size of an image without damaging the overall structure or losing important information in the image.

INDEX TERMS Image retargeting, seam carving, CNN, image resize.

I. INTRODUCTION

With the proliferation of multimedia devices with various screen sizes, image resolutions have become increasingly diverse. The process of resizing has become essential in every device, to display images correctly. In the conventional resizing methods, the overall image size is reduced. When the size of the entire image is reduced, the resolution also decreases. The common resizing techniques include bilinear and bicubic techniques in which changes are made linearly. However, most of the widely used linear methods have high probability of distorting the content or losing important information. To overcome such a drawback, the interested area/part in an image can be cropped. The cropping method is the function of cutting an area of appropriate ratio or desired size, in an image, by using a quadrangular box, thereby removing the unimportant parts. However, it has the disadvantage that information is lost or distorted when there are multiple important parts in an image.

Recently, studies have been carried out on retargeting methods that decrease the resolution of an image while preserving the content. These methods obtain and output an image suitable for a particular display, without distorting major objects while resizing the image. Several studies have been conducted on the method of seam carving, which is a method of adjusting an image size based on energy biases or features. It defines a continuous one-dimensional seam in the direction that minimizes damage to objects, in an image constructed with an isotropic seam. By removing the seam sequentially, the image size can be adjusted stably. Content-aware image resizing, proposed by Avidan *et al.* [1], was the first paper on seam carving. Thereafter, this technique was discussed in many studies and several modifications were made. For example, using a Sobel mask, segmentation, HoG, entropy, etc., features of images were detected and a feature map was created. Then, an energy pattern was obtained from the feature map created and the original image. Finally, dynamic programming method was used for reducing or stretching an area based on the path along which the energy was minimized. In this case, since the image's features and direction were considered, the image size could be reduced without damaging important information in the image.

The effectiveness of seam carving lies in the robust calculation of energy bias. When detecting the bias, a threshold value is set and, depending on this threshold value, the intensity of the energy map varies. However, by fixing the threshold

FIGURE 1. Results of the scale and seam carving methods and our proposed algorithm.

value, a problem arises in that it cannot respond to diverse images. The conventional method calculates an energy map using a 3×3 differentiation filter. However, during the calculation process, the energy intensity of the energy map varies depending on the image. Therefore, using an encoder–decoder convolution neural network (E–D CNN), we propose a method for creating an energy map according to a learning model. Our proposed deep energy map creation network can create energy maps stably for simple/complex images by learning various images. Moreover, a much better result can be obtained for seam carving.

In this paper, a deep convolution neural network (DCNN) is implemented using the structures of an encoder and a decoder. An image retargeting system is constructed, which considers the content and objects of images when decreasing or increasing the image size. Our final goals can be summarized in three main points. First, using the neural network, a deep energy map creation model is proposed. Second, the energy map is applied to the retargeting system that can adjust the size of an image. Third, a reliable evaluation method is used for evaluating the improved retargeting system.

The remainder of this paper is organized as follows. In Section 2, image retargeting is introduced and its characteristics are described. The related studies on image resizing are explained and studies using seam carving are introduced. In Section 3, the encoder–decoder structure used in our proposed DCNN is explained, and the framework that combines seam carving with the deep energy map created from the DCNN is described in detail. In Section 4, the test environment, dataset, evaluation indexes, and evaluation results are described. The conclusions and future work are provided in the last section.

II. RELATED WORK

In this section, the previous studies on image processing and image retargeting using the CNN are reviewed. A method that uses the DCNN and a method that improves image retargeting are introduced.

A. DEEP CONVOLUTION NEURAL NETWORK

Recently, with the advances in artificial intelligence, signal processing methods using DCNN have been introduced. Typical examples include super resolution [2], image artifact removal [3], deblurring [4], colorization [5], image in-painting [6], and image matting [7]. All the aforementioned method process or restore images using the learning model of CNN.

The DCNN technology that is to be merged in our proposed study is of the edge detection field and can detect the foreground, background, or an object's outlines from an image. An edge has the characteristic that it shows the boundary of an area in an image, and it indicates a discontinuity in pixel brightness. Furthermore, it corresponds to the contour of an object and contains much significant information. In general, it is detected as the approximate value of a differential operation by using the difference between adjacent pixels. A large gradient in the differentiation value is due to a large change in the brightness. Therefore, it can be inferred that a boundary of the object exists at that point. Operators such as the Sobel, Prewitt, and Robert operators are the most widely used for the first-order derivative detection and operators such as the Laplacian, Laplacian of Gaussian, and Canny are widely used for the second-order derivative detection. These edge detection algorithms use classical image processing [8]–[10]. Recently, using machine learning, several studies were conducted on the accuracy and speed of edge detection. Xiaofeng *et al.* [10] combined the proven concept of oriented gradients with powerful representations that were automatically learned through sparse coding. Sparse code gradients (SCG) performed significantly better than the hand-designed features that were in use for a decade. Dollar *et al.* [11] proposed a generalized structured learning method for learning structured random decision forests that robustly used structured labels to select splits in the trees. As for the edge detection rate, the result was better, by a narrow margin, than that of the method using SCG. By detecting edges through learning, edges of uniform intensity can be obtained. Several studies have actively tried to find edges using deep learning. Xie *et al.* [12] developed an end-to-end edge detection system, holistically-nested edge detection (HED), which automatically learnt the type of rich hierarchical features that were crucial if we were to approach the human ability to resolve ambiguity in natural image edges and for object boundary detection. The most noticeable study is the Richer Convolutional Features for Edge Detection proposed by Liu *et al.* [13] Using deep learning, edges were detected in a manner most similar to the detection by human eyes. The evaluation was conducted with a BSDS

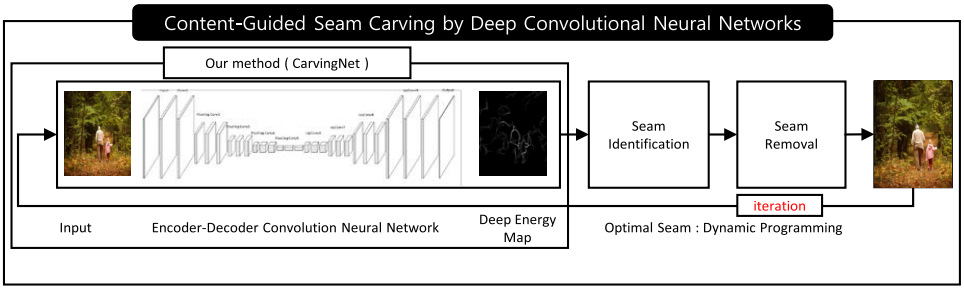


FIGURE 2. Image retargeting system using the proposed deep convolutional neural network (DCNN) and seam optimization.

500 dataset [30]. The evaluation results were measured with an accuracy of 0.806 for the ODS F-measure. While there have been many studies conducted, there are few research results on image size control using deep running. As mentioned above, we hope that our image retargeting method using DCNN will contribute to the development of the image processing field.

B. IMAGE RETARGETING

Ever since retargeting was proposed by Avidan *et al.* [1] in 2007, studies pertaining to its performance improvement have been conducted. Many existing algorithms determine the importance of pixels in an image and decrease or increase the image size through arithmetic operations. Typically, retargeting methods are divided into two categories: discrete methods that remove or move pixels in an image and continuous methods that use a quadrangular mesh. When the result of an arithmetic operation can be expressed as an integer, it is called discrete.

The Discrete algorithms proposed previously include seam carving algorithm [1], [14]–[16], [18]–[21], shift map [17], [18], video retargeting [22]–[24], and multiple seams [15]. All these algorithms repeatedly remove or add one-dimensional seams that are not important in an image. Avidan *et al.* [1] introduced the concept of seam carving and adjusted the aspect ratios through dynamic programming. Goferman *et al.* [18] reduced the size of an image by creating a shift map from an image. The shift map was computed by optimal graph labeling, wherein a node in the graph corresponded to a pixel in the output image. Rubinstein *et al.* [16] proposed seam carving for videos, which worked by enforcing temporal coherence of content-aware video warping by solving a global optimization problem over the entire video cube. Han *et al.* [15] proposed a novel method to find multiple seams simultaneously with global optimality for image resizing, incorporating both region smoothness and seam shape prior to using a three-dimensional graphical–theoretic approach.

The Continuous methods [25]–[28] isolate a certain section and perform an arithmetic operation stochastically on the pixels included in that section, because they cannot count. The methods presented recently include scale-and-stretch, feature-aware texturing, mesh parametrization, and triangular mesh. Wang *et al.* [25] presented a "scale-and-stretch" resizing method that simultaneously warped the input image in both directions. The scale-and-stretch method introduced edge bending energy, which could effectively prevent local bending of features, but could not prevent the distortion to the global structure. Gal *et al.* [26] described a feature-aware texturing method that warped an original image into an arbitrary shape while preserving the shapes of user-specified features by constraining the warping to be a similarity transformation. Guo *et al.* [27] suggested a mesh representation based on image structures to preserve the shape of an input image. Jin *et al.* [28] presented a novel approach for content-aware image resizing by using a triangular mesh. They proposed a triangular mesh over an original image that was consistent with the existing image features, which included sharp edges, feature curves, and image boundaries. Then, a global nonhomogeneous scale optimization was performed to warp the triangular mesh. The warped triangular mesh and the horizontal and vertical scales of all triangles were obtained simultaneously by a quadratic optimization, which could be performed by solving a sparse linear system.

III. NETWORK ARCHITECTURE

In this section, we introduce the content-guided seam-carving technique that uses the encoder–decoder CNN (E–D CNN). Figure 2 shows the framework proposed by us. The framework consists of three main stages: the first stage is the E–D CNN stage, which is followed by the seam identification stage, and the last stage is the seam removal stage.

A. ENERGY MAP GENERATOR USING E-D CNN

Using the E–D CNN, we propose the energy map creation. Figure 3 shows the framework for energy map creation and seam carving. Its architecture is similar to the network architecture of the AutoEncoder. The difference from the AutoEncoder is that a convolution product layer is used in the encoder stage whereas a deconvolution product layer, which follows the reverse process, is used in the decoder stage. Furthermore, all output values of the layers constituting the encoder stage are used in the decoder stage. Therefore, it is an appropriate model for conversion to a certain shape

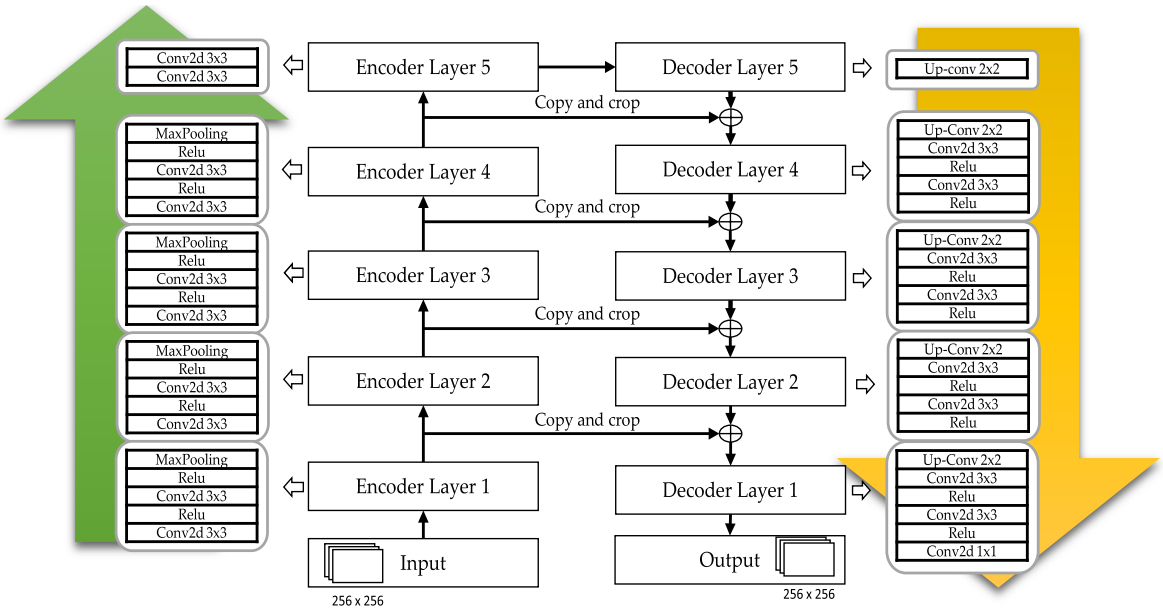


FIGURE 3. Encoder–decoder convolutional neural network (E–D CNN) Architecture: (a) The green box is the encoder and (b) the yellow box is the decoder.

such as energy map creation. The E–D CNN follows an end-to-end methodology. As shown in Figure 3, the architecture consists of downscaling and upscaling stages, and in the encoder stage, max pooling and convolution product are used. Max pooling divides an input image into 2×2 images, and outputs the largest value from each image. For a 2×2 max pool operation including stride 2, the image size is decreased to $\frac{1}{2}$.

The decoder stage performs the reverse function of the encoder stage. If a feature map is created in the $\frac{1}{2}$ scale in the encoder, the feature map are increased twice by 2×2 up-convolution in the decoder. The decoder feature map and encoder feature map output during the encoder process are combined and the deconvolution product is calculated. The combined feature map is changed to a large image as two convolution product processes and Relu are performed. Furthermore, we want the predictions to reflect the probability of a pixel being a seam or a minimum path; therefore, we use a sigmoid activation function in the last layer.

Table 1 describes the E–D CNN in detail. A three-channel (RGB) image of resolution 256 (width) $\times$ 256 (height) is received as the input image. The input image goes through the network consisting of seven unit levels. At the output, an energy map of 256×256 pixels and one-channel resolution is obtained. The core of E–D CNN is a network structure that creates strong energy at the important parts in an image and weak energy at the unimportant parts. The proposed network creates stochastically robust deep energy maps. We made a training model by using the learning dataset of BSDS500. It had an end-to-end structure and a color image of resolution 256×256 was received as the input. When creating the training model, Nvidia Geforce 1080 was used in the hardware. The batch size used for training was 1. For the optimization algorithm, an Adam optimizer was chosen, and the learning rate was set to 0.0001. Using the GeForce 1080, it took 240 minutes to complete 200 epochs.

TABLE 1. Structures of Layer, Filter, and Stride are shown in the Encoder-Decoder CNN, and the output size changes are shown according to the filter size.

	Unit level	Conv layer	Filter	Stride	Output size
Input					$256 \times 256 \times 3$
Encoding	Level 1	Conv 1	$3 \times 3/32$	1	$256 \times 256 \times 32$
		Conv 1	$3 \times 3/32$	1	$256 \times 256 \times 32$
	Level 2	Conv 2	$3 \times 3/64$	2	$128 \times 128 \times 64$
		Conv 2	$3 \times 3/64$	1	$128 \times 128 \times 64$
	Level 3	Conv 3	$3 \times 3/128$	2	$64 \times 64 \times 128$
		Conv 3	$3 \times 3/128$	1	$64 \times 64 \times 128$
	Level 4	Conv 4	$3 \times 3/256$	2	$32 \times 32 \times 256$
		Conv 4	$3 \times 3/256$	1	$32 \times 32 \times 256$
Bridge	Level 5	Conv 5	$3 \times 3/512$	2	$16 \times 16 \times 512$
		Conv 5	$3 \times 3/512$	1	$16 \times 16 \times 512$
Decoding	Level 6	De-conv 6	$3 \times 3/256$	1	$32 \times 32 \times 256$
		De-Conv 6	$3 \times 3/256$	2	$32 \times 32 \times 256$
	Level 7	De-conv 7	$3 \times 3/128$	1	$64 \times 64 \times 128$
		De-conv 7	$3 \times 3/128$	2	$64 \times 64 \times 128$
	Level 8	De-conv 8	$3 \times 3/64$	1	$128 \times 128 \times 64$
		De-conv 8	$3 \times 3/64$	2	$128 \times 128 \times 64$
	Level 9	De-conv 9	$3 \times 3/32$	1	$256 \times 256 \times 32$
		De-conv 9	$3 \times 3/32$	2	$256 \times 256 \times 32$
Output		De-conv 10	1×1	1	$256 \times 256 \times 1$

B. CONTENT-GUIDED SEAM CARVING

1) GENERATION OF ENERGY MAP USING SEAM CARVING

Seam carving extracts a seam that has the lowest energy change. By adding or removing one-dimensional input information on the coordinates of the corresponding seam, the image can be resized. The seam is connected horizontally or vertically in the image, and it is a line where the row or column is connected by only one pixel. The traditional seam carving algorithm creates a gradient-based energy map and then determines important areas of the image. Equation 1 is the equation for a gradient-based energy map. In the

calculation of the gradient energy graphs, if I is assumed to be a given image, and the size is given by width Ã- height, the gradient energy function of the image is defined as:

$$Energy(I) = \left|\frac{\partial}{\partial x}I\right| + \left|\frac{\partial}{\partial y}I\right| \quad (1)$$

The essence of Equation 1 is that the sum of gradients for the x-direction and y-direction is the energy value of the pixel. If the pixel energy map is large, it indicates important information that must be maintained. Conversely, when the energy is less, there is a high probability that it is not important. Therefore, the energy map shows the importance of the pixel in an image.

2) GENERATION OF ENERGY MAP USING E-D CNN MODEL

Although the deep energy map using the E–D CNN proposed in this paper has the same functions and goals as the conventional energy map, it was proposed to produce a higher performance. What is important here is cross entropy. Cross entropy is defined to show the difference between two probability distributions. More specifically, cross-correlation (CC) is used when showing a correlation between two functions whereas cross entropy (CE) shows the distance between two probability distributions. Therefore, in the E–D CNN, the final cross entropy loss function outputs an energy map by combining the final feature map and Sigmoid function. The cost function is expressed as shown in Equation 2.

$$C = -\frac{1}{n}\sum_{x}[y\ln f(s_1) + (1-y)\ln(1-f(s_1))] \quad (2)$$

where $-[y\ln f(s_1) + (1-y)\ln(1-f(s_1))]$ is the binary entropy, n is the number of training data, and y is the output value required for the training data. $f()$ is the sigmoid function. so the gradient respect to the each score s_i in s will only depend on the loss given by its binary problem. The cross-entropy function has two properties: non-negativity and cross-entropy $= 0$. Regarding the non-negativity, all terms inside the sigma are always negative numbers, and the terms outside sigma have a value of 0 or greater, because of the negative sign. In addition, to elaborate on the number of two cases, when $y = 0$, if s_1 is close to 0, the cross-entropy term will be close to 0 because $ln(1)$ is close to 0. When $y = 1$, if s_1 is close to 1, the cross-entropy term will be close to 0 because $ln(1)$ is close to 0. We substitute $s_1 = \sigma(z)$ in Equation 2. Then, if w is differentiated using the chain rule, since the differential value of $\ln(x)$ is $\frac{1}{x}$, it can be shown as in Equation 3.

$$\begin{aligned}\frac{\partial C}{\partial w_j} &= -\frac{1}{n}\sum_{x}\left(\frac{y}{\sigma(z)} - \frac{(1-y)}{1-\sigma(z)}\right)\frac{\partial\sigma}{\partial w_j} \\ &= -\frac{1}{n}\sum_{x}\left(\frac{y}{\sigma(z)} - \frac{(1-y)}{1-\sigma(z)}\right)\sigma'(z)x_j \end{aligned} \quad (3)$$

If Equation 3, which is a sigmoid function, is substituted in Equation 4,

$$\sigma'(z) = \sigma(z)(1-\sigma(z)) \quad (4)$$

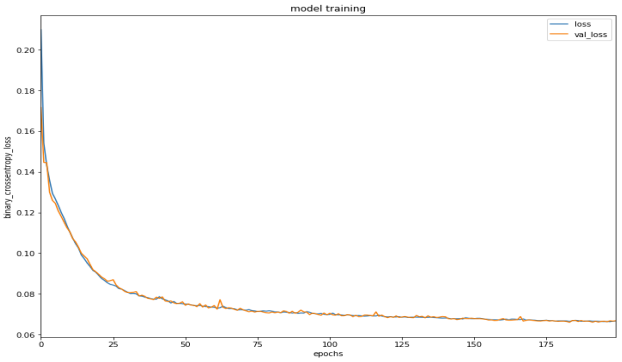


FIGURE 4. Loss of training network.

it can be summarized as in Equation (5).

$$\frac{\partial C}{\partial w_j} = \frac{1}{n}\sum_{x} x_j(\sigma(z) - y) \quad (5)$$

In Equation 5, the changing rate of weight is determined by $\sigma(z) - y$. As the difference of the two becomes larger, the gradient of change becomes larger. Finally, in all training images, if $\sigma(z) = y$ is substituted in the cross-entropy function, then C of Equation 2 will be minimized. We have designed it to be output by the inputs and biases coming into the corresponding neuron, and we called it the deep energy map.

In Figure 2, the seam identification stage is the function that is processed after the creation of the energy map [1]. The method of finding the smallest or largest energy pattern by comparing the difference between energy pixels in the deep energy map is called seam identification. Equation 6 finds the difference of values between all neighbor pixels existing in the one-dimensional vertical seam direction, thereby, finding the pattern in the direction where the difference is the smallest.

$$\begin{aligned} &S^x = \{S_i^x\}_{i=1}^n = \{(x(i), i)\}_{i=1}^n, \\ &where \quad \forall i, \quad |x(i) - x(i-1)| \le 1 \end{aligned} \quad (6)$$

Similarly, a horizontal seam is defined as:

$$\begin{aligned} &S^y = \{S_j^y\}_{j=1}^m = \{(y(j), j)\}_{j=1}^m, \\ &where \quad \forall j, \quad |y(j) - y(j-1)| \le 1 \end{aligned} \quad (7)$$

The vertical seam indicates a one-dimensional path coming down from the top to the bottom and the energy sum of the one-dimensional path is given by S. Similarly, a horizontal seam indicates a one-dimensional path flowing from the left to the right, and Equation 7 expresses the sum of one-dimensional energies in the horizontal direction. The seam where the cumulative sum of seams is a minimum becomes the optimum seam. Since we process the vertical seam and horizontal seam in the same manner, the explanation is provided focusing on the vertical seam. In order to find this seam, we first define the cumulative minimum energy map M for

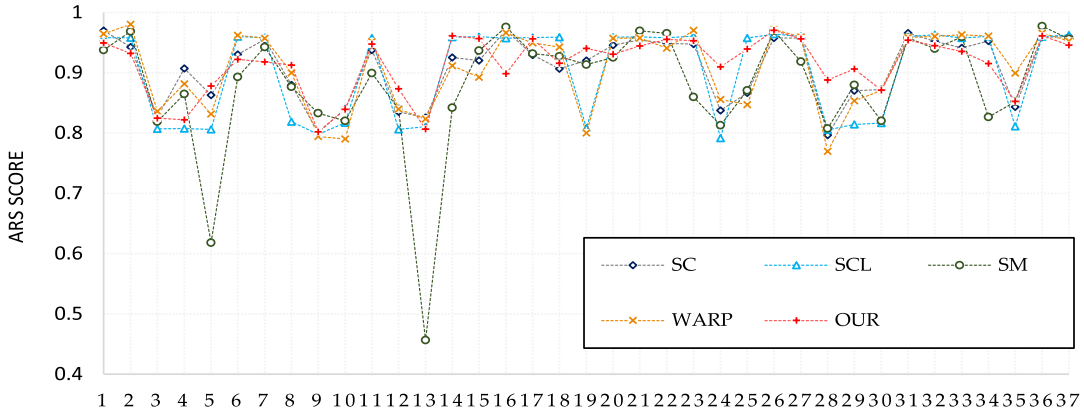


FIGURE 5. Graph of aspect ratio similarity metric results for the test set of 37 images (seam carving (SC), scaling (SCL), shift map (SM), non-homogeneous warping (WARP), proposed method).

the second row to the last row as follows:

$$M(i,j) = e(i,j) + min(M(i-1,j-1), M(i-1,j), \quad M(i-1+j)) \tag{8}$$

As shown by in Equation 8, the pixel that is a the minimum in the a row and column is selected as the initial position. Then, from among three rows that are adjacent to the pixel right mentioned above, a minimum value is obtained.

Finally, the seam removal, which is the last stage in Figure 2, removes the seam where the energy is a minimum. By repeating the above process, the image size can be reduced effectively.

C. LOSS FUNCTION

We used the Dice coefficient itself for the loss function. Specifically, Dice coefficient is defined as follows:

$$DiceSimilarity(EM, P) = \frac{2 \cdot |EM \cap P|}{|EM| + |P|} \tag{9}$$

where EM(Edge Mask) is the predicted set of pixels and P is the ground truth. A higher dice coefficient is better. The EM and P is the output of a sigmoid function (0-1). Since the denominator is constant, the only way to maximize this metric is to increase overlap between EM and P.

Figure 4 Graph showing the dice coefficient. The dice coefficient in figure 4 is higher than 0.07. It can be seen that the Dice for training is still in a decreasing trend.

IV. EXPERIMENTS

A. SETUP

The hardware configuration used in the experiments is a desktop computer (Intel® Core(TM) i7-7700 CPU @ 3.6 GHz, 24.0 GB RAM), and the seam carving algorithm is implemented with Python version 3.6. The CNN is directly downloaded from Tensorflow and Keras (https://www.tensorflow.org, https://keras.io)

B. ASPECT RATIO SIMILARITY METRIC(ARS)

The aspect ratio similarity proposed by Zhang *et al.* [29] is a framework that creates resampling grids and integrates forward resampling. ARS is an effective method that can estimate the geometric transition for the relationship of images. More specifically, the visual quality inside a local block is evaluated by an importance pooling strategy. We perform an objective evaluation using ARS. First, the ARS measures the similarity as shown in Equation 10. w_{ret} is the maximum width in the block and h_{ret} is the maximum height in the block. The height and width change ratios can be denoted as $r_w = \frac{w_{rt}}{N}$ and $r_h = \frac{h_{rt}}{N}$ and $r_h = \frac{h_{rt}}{N}$ and the mean ratio $u_r = \frac{r_w + r_h}{2}$ denotes the absolute block size change. The similarity score of the block pair is formulated as follows:

$$S = \left[\frac{2 \cdot r_w \cdot r_h + C}{r_w^2 + r_h^2 + C}\right] \cdot \left[e^{-\alpha(\mu_r - 1)^2}\right] \tag{10}$$

Here, C is a variable for increasing the stability to make provision for the case of dividing by 0. $e^{-\alpha}$ is a penalty adjusting parameter to deal with information loss caused by image distortion. Finally, using an index from 0 to 1, S shows the level of image loss or visual distortion in a block. When the index is close to 1, the image quality in the block is high, and if S is close to 0, the information loss in the block is large. Next, ARS can be shown as in Equation 11

$$ARS = \sum_m \sum_n S_{mn} \cdot V_{mn} \Big/ \sum_m \sum_n V_{mn} \tag{11}$$

Here, m and n are block indices of the original image, and Smn is the similarity score of the respective indices. Vmn is the index variable of the visual importance map. Finally, the ARS is evaluated with an index between 0 and 1.

C. EXPERIMENTAL RESULTS

Seam carving has a characteristic that a one-dimensional seam is removed by focusing on the seam when reducing an image. We evaluate the loss level of an image by changing to various resolutions, but when reduced to 20% or less,

FIGURE 6. Resulting images of five algorithms for the images of retarget me dataset. Result images of (a) SCL, (b) seam carving, (c) shift map, (d) WARP, and (e) proposed method. Here, the number 1 denotes the result of reducing to 25% size and 2 denotes the result of reducing to 50%.

TABLE 2. Evaluation result of aspect ratio similarity for 37 test images (average and dispersion).

		SC[1]	SCL	SM	WARP[24]	Our Method
ARS	Average	0.909	0.902	0.880	0.908	**0.913**
	Std	0.053	0.074	0.101	0.064	**0.047**

it is difficult to find anything special about the algorithm. Therefore, focusing on the vertical seam, we evaluate the level of loss by reducing the retarget me dataset [31] images to 25% (23 frames) and 50% (14 frames).

In Figure5 we define the seam carving (SC), scaling (SCL), shift map (SM), non-homogeneous warping (WARP), and the proposed method. In Table 2, five algorithms are evaluated in terms of the ARS. The average score of our proposed method is 0.033 higher than that of SM, and the standard deviation is 0.054 lower. Compared to the SC, the average score is 0.004 higher. This signifies that, with respect to the distortion or loss level of images, the proposed method is excellent, and its deviation is not severe for the images (lines/edges, face/people, foreground objects, texture, geometric structures, and symmetry). In general, seam carving carves the sides where the energy of seams is small in an image whereas the method proposed by us provides a result wherein the distortion is less and information loss is decreased because the energy is maintained at important places in the image and the places where the energy is relatively small are carved.

V. CONCLUSION

The content-guided seam carving by DCNN proposed in this paper could change the size of an image to a required size while preserving the feature properties well. Conventional seam carving created an energy map by using differential arithmetic operations to find a seam where the energy was the minimum. However, in such an energy map, the intensities of the edges were not uniform. Because of this, the result was not good owing to the cutoff phenomenon and background distortion. To overcome these drawbacks, we created an energy map in which the complexity of color and texture were robust and the energy density was uniform, by training the features of images and boundary data. Furthermore, using the ARS evaluation method, an objective evaluation was performed. Compared to the differential energy map, the proposed deep energy map method showed higher performance. Particularly, the line segments and boundaries of image features were well maintained. Furthermore, the energy map conversion was robust to various images, whereby, relatively stable energy maps were created. Therefore, by using deep energy maps, the user's desired resizing results could be obtained. Thus, the limitations of conventional technology in which the results calculated automatically by a computer could not always satisfy the user, could be overcome. In the future, our system will be able to overcome the limitations associated with the resolutions by improving the method using DCNN and will be able to improve the speed by increasing the efficiency of the arithmetic operations.

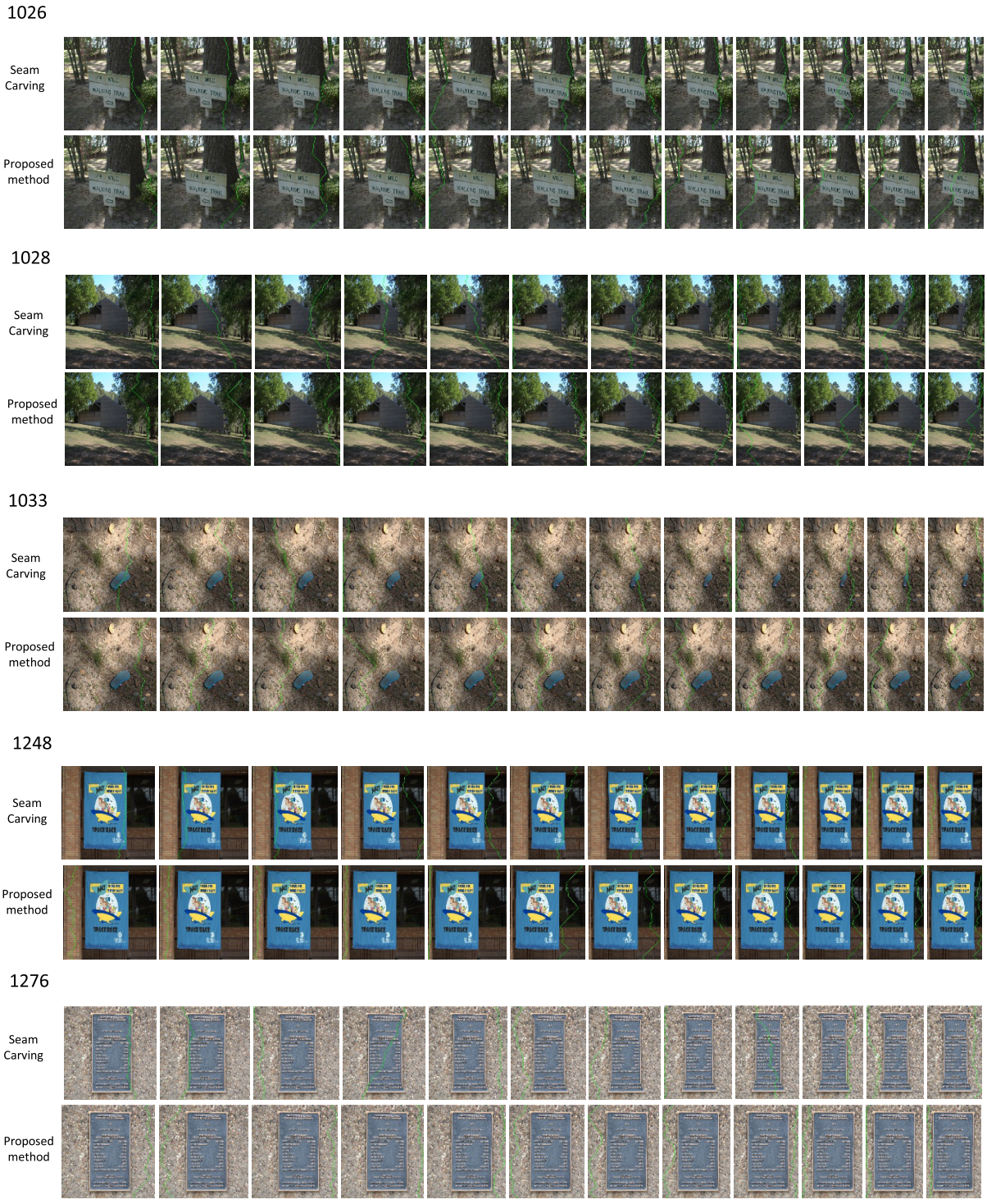


FIGURE 7. Graph of aspect ratio similarity metric results for the test set of 37 images (seam carving (SC), scaling (SCL), shift map (SM), non-homogeneous warping (WARP), proposed method).

APPENDIX FOR "CONTENT-GUIDED SEAM CARVING USING DEEP CONVOLUTION NEURAL NETWORK"

See Figure 7.

REFERENCES

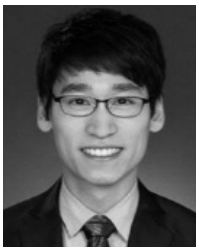
[1] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *Trans. Graph.*, vol. 26, no. 3, Jul. 2007, Art. no. 10.

[2] Y. Li, W. Dong, X. Xie, G. Shi, X. Li, and D. Xu, "Learning parametric sparse models for image super-resolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4664–4672.

[3] L. Xu, J. S. Ran, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1790–1798.

[4] T. M. Nimisha, A. K. Singh, and A. N. Rajagopalan, "Blur-invariant deep learning for blind-deblurring," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4760–4770.

[5] Z. Cheng, Q. Yang, and B. Sheng, "Deep colorization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 415–423.
[6] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 1, Dec. 2012, pp. 341–349.
[7] N. Xu, B. Price, S. Cohen, and T. Huang, "Deep image matting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2970–2979.
[8] Y. Ganin and V. Lempitsky, "N^4-fields: Neural network nearest neighbor fields for image transforms," in *Proc. ACCV*, Jun. 2014, pp. 536–551.
[9] J. Kivinen, C. Williams, and N. Heess, "Visual boundary prediction: A deep neural prediction network and quality dissection," in *Proc. 17th Int. Conf. Artif. Intell. Statist.*, 2014, pp. 512–521.
[10] R. Xiaofeng, and L. Bo, "Discriminatively trained sparse code gradients for contour detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 584–592.
[11] P. Dollar and C. L. Zitnick, "Structured forests for fast edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1841–1848.
[12] S. Xie, and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1395–1403.
[13] Y. Liu *et al.*, "Richer convolutional features for edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017.
[14] A. Shamir and S. Avidan, "Seam carving for media retargeting," *Commun. ACM*, vol. 52, no. 1, pp. 77–85, Jan. 2009.
[15] D. Han, M. Sonka, J. Bayouth, and X. Wu, "Optimal multiple-seams search for image resizing with smoothness and shape prior," *Visual Comput.*, vol. 26, nos. 6–8, pp. 749–759, Jun. 2010.
[16] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *Trans. Graph.*, vol. 27, no. 3, p. 16, 2008.
[17] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Oct. 2009, pp. 151–158.
[18] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.
[19] I. Izumi, "Gradient-based global features for seam carving," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, p. 27, 2016.
[20] T. Yin, G. Yang, L. Li, D. Zhang, and X. Sun, "Detecting seam carving based image resizing using local binary patterns," *Comput. Secur.*, vol. 55, pp. 130–141, 2015.
[21] D. Cho, J. Park, T.-H. Oh, Y.-W. Tai, and I. S. Kweon, "Weakly- and self-supervised learning for content-aware deep image retargeting," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4568–4577.
[22] S. D. Desai, M. Bhille, and N. D. Hiremath, "Content-aware video retargeting by seam carving," in *Proc. 5th Int. Conf. Frontiers Intell. Comput. Theory Appl.* Singapore: Springer, Mar. 2017, pp. 147–157.
[23] B. Guthier, J. Kiess, S. Kopf, and W. Effelsberg, "Seam carving for stereoscopic video," in *Proc. IVMSP*, Jun. 2013, pp. 1–4.
[24] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Proc. 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–6.
[25] Y.-S. Wang, C. L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *Trans. Graph.*. vol. 27. no. 5, Dec. 2008, Art. no. 118.
[26] R. Gal, O. Sorkine, and D. Cohen-Or, "Feature-aware texturing," in *Proc. 17th Eurograph. Symp. Rendering*, Jun. 2006, pp. 297–303.
[27] Y. Guo, F. Liu, J. Shi, Z.-H. Zhou, and M. Gleicher, "Image retargeting using mesh parametrization," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 856–867, Aug. 2009.
[28] Y. Jin, L. Liu, and Q. Wu, "Nonhomogeneous scaling optimization for realtime image resizing," *Visual Comput.*, vol. 26, nos. 6–8, pp. 769–778, Jun. 2010
[29] Y. Zhang, W. Lin, X. Zhang, Y. Fang, and L. Li, "Aspect ratio similarity (ARS) for image retargeting quality assessment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2016, pp. 1080–1084.
[30] *Set, Berkeley Segmentation Data Benchmarks 500 (BSDS500)*. [Online]. Available: http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html
[31] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *Trans. Graph.*, vol. 29, no. 6, Dec. 2010, Art. no. 160.



EUNGYEOL SONG received the B.S. degree from the Department of Electronic and Electrical Engineering, Dankook University, in 2012. He is currently pursuing the Ph.D. degree in electrical and electronic engineering with Yonsei University, Seoul, South Korea. His current research interests include hand tracking, 3-D reconstruction, and machine learning.

MINKYU LEE received the B.S. degree from the Department of Computer Engineering, Kwangwoon University, in 2013. He is currently pursuing the Ph.D. degree in electrical and electronic engineering with Yonsei University, Seoul, South Korea. His current research interests include camera calibration and visual tracking.

SANGYOUN LEE (M'04) received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 1987 and 1989, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1999. He is currently a Professor and the Head of electrical and electronic engineering at the Graduate School, and the Head of the Image and Video Pattern Recognition Laboratory, Yonsei University. His research interests include all aspects of computer vision, with a special focus on pattern recognition for face detection and recognition, advanced driver assistance systems, and video codecs.