Image Enhancement for Attractiveness Computing

Toshihiko YAMASAKI

yamasaki@cvm.t.u-tokyo.ac.jp
Department of Information and Communication Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
Keywords: Artificial Intelligence (AI), Image Design, Creativity

ABSTRACT

Deep Neural Networks (DNNs) and big data on the Internet have enabled us to achieve better image processing and editing than ever such as retouching, inpainting, style transfer, and so on. Our research group has been working on attractiveness prediction, reasoning, and even enhancement for images and videos, which we call "attractiveness computing." Image processing in attractiveness computing also pay attention to image quality, not in terms of signal-to-noise ratios, but in terms of higher-level human affects such as impressiveness, instagrammability, memorability, clickability, and so on. Analyzing such attractiveness was usually done by experienced professionals but we have experimentally revealed that properly designed algorithms can perform as well as skilled professionals. In this paper, we introduce some of our representative works and possible applications of our attractiveness computing for image enhancement.

1 Introduction

It has been almost a decade since Deep Neural Networks (DNNs) gave a big impact to our research community by demonstrating overwhelming performance in image recognition in the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) [1]. However, image recognition tasks are relatively easier to tackle because ground truths can be strictly defined, and the performance can be objectively evaluated.

On the other hand, our research group is interested in more subjective problems: why some of the image/video contents are more attractive, impressive, appealing, etc., and how we can make the contents more attractive. Usually, it is thought that making images and videos more attractive needs sense and skill and therefore it is hard for ordinary people to achieve it. Our research group call a task of predicting, reasoning, and enhancing the attractiveness of multimedia content as "attractiveness computing," and has been doing research such as presentation analysis [1]-[4], popularity analysis in social network [5]-[7], and so on. In this paper, some of our representative works on image processing and image quality enhancement are introduced.

2 Ultra-Fast Image Retargeting

Image retargeting, which is a task of adjusting input images to arbitrary sizes, is fundamental image processing to fit the input image to the given display sizes. However, image retargeting is generally computationally very expensive. We have proposed some optimization methods for ultra-fast image retargeting for 2D images [8] and 3D volume data [9].

In one of the best-performing methods for 2D images called MULTIOP, the multiple retargeting operators were combined and retargeted images at each stage were generated to find the optimal sequence of operators that minimized the distance between original and retargeted images. The limitation of this method is in its tremendous processing time, which severely prohibits its practical use. Therefore, we proposed a method of predicting the optimal retargeting operator for each step using a reinforcement learning agent [8]. The technical contributions are as follows. Firstly, we proposed a reward based on self-play, which will be insensitive to the large variance in the content-dependent distance measured in MULTIOP. Secondly, we proposed to dynamically change the loss weight for each action to prevent the algorithm from falling into a local optimum and from choosing only the most frequently used operator in its training. Our experiments showed that our method achieved multi-operator image retargeting with less processing time by three orders of magnitude and the same quality as the original multi-operator-based method, which was the best-performing algorithm in retargeting tasks. The architecture of our proposed method is shown in Figure 1 and some retargeting results along with generated operator sequences are shown in Figure 2.

In volume seam carving, i.e., seam carving for 3D cost volume, an optimal seam surface can be derived by graph cuts and it was the only solution. However, it is not suitable for practical use because it incurs a heavy computational load. We proposed a multipass dynamic programming (DP)-based approach for volume seam carving, which reduces computational time and memory consumption while maintaining a similar image quality as that of graph cuts. As can be seen in Figure 3, our proposed method is very efficient while achieving visually pleasing results.

[1] https://image-net.org/challenges/LSVRC/2012/

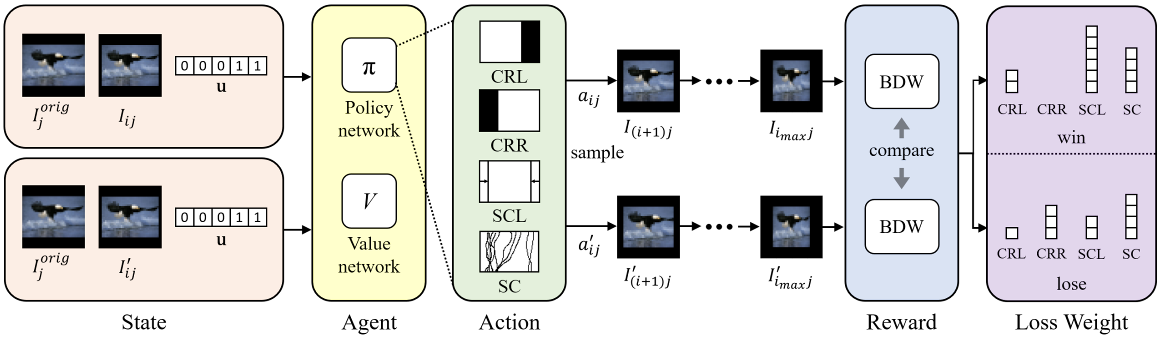

Figure 1 Illustration of the self-play reinforcement-learning-model architecture. The agent repeats receiving of the observation composed of the original image, the current retargeted image, a one-hot vector representing the number of steps to the end of the episode, and sampling two actions of the self and the opponent according to the policy output. Each current retargeted image is updated by the retargeting operator corresponding to the sampled action. At the end of the episode, the agent receives a reward based on the victory or defeat of the BDW score. The loss weight of the value network and the policy network are changed according to the number of times each action is selected in case of victory or defeat. CRL and CRR: cropping left/right, SCL: scaling, and SC: seam carving.

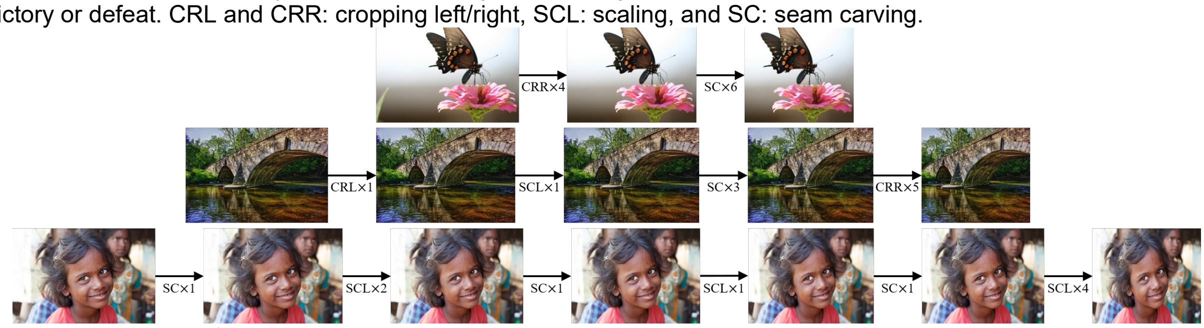

Figure 2 Examples of sequential actions selected by the agent. The symbol below the arrow indicates the applied retargeting operator and the number of times it is applied. (©Lindsey Turner, Ted Van Pelt, Nithi Anand)

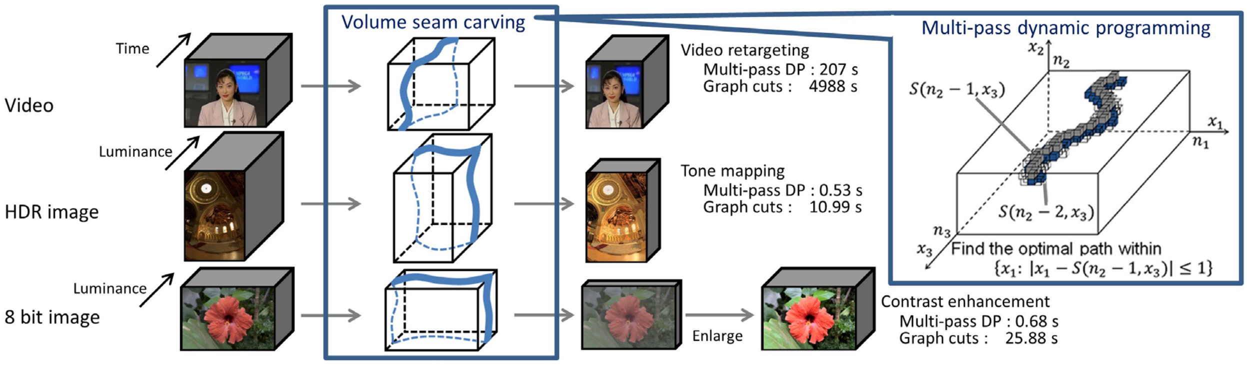

Figure 3 Multipass DP for volume seam carving proposed to reduce computation time and memory consumption. Three example applications of multipass DP are presented.

3 Reinforcement Learning for Image Processing

Unpaired image-to-image translation, a task of learning the mapping from a source domain to a target domain in the absence of paired images, has achieved significant results owing to the recent development of neural-network-based generative adversarial networks (GANs). Instead of simply generating images with a neural network, we proposed a method to translate images with image editing software such as Adobe Photoshop for the following three benefits: translated images have no artifacts, the same translation can be applied to larger images, and the translation is interpretable [10]. To incorporate image editing software into a GAN, we proposed a reinforcement learning framework where the generator works as the agent that selects the software's parameters and is rewarded when it fools the discriminator (Figure 4). Our framework can use high-quality non-differentiable filters present in image editing software, which enables translation with high performance. Our experimental results shown in Figure 5 demonstrate that the proposed method achieves good performance.

We also proposed reinforcement learning with pixel-wise rewards (pixelRL) for image processing [11][12]. After the introduction of the deep Q-network, deep RL has

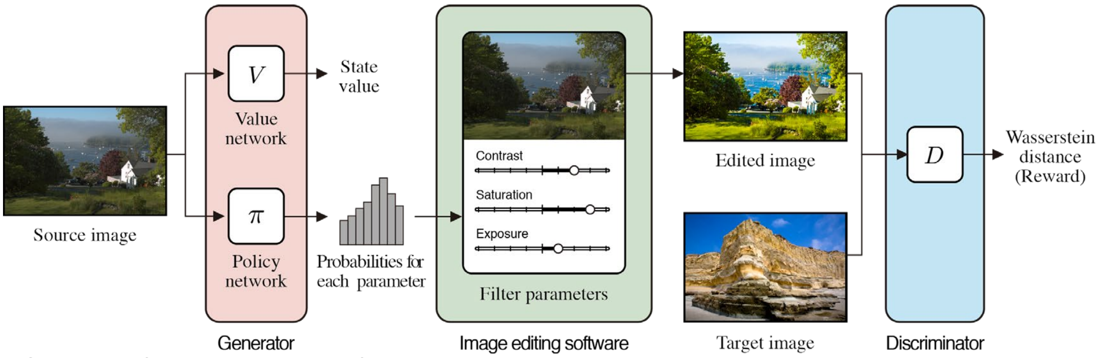

Figure 4 Overview of our method. In our framework, the generator is trained with RL to control image editing software, and the output of the discriminator is used as the reward.

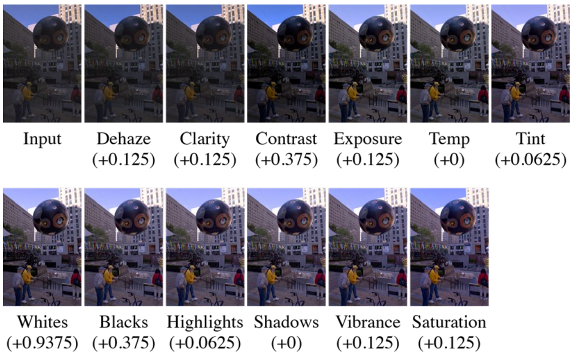

Figure 5 How the image filters are applied for enhancing the image quality.

been achieving great success. However, the applications of deep reinforcement learning (RL) for image processing are still limited. Therefore, we extended deep RL to pixelRL for various image processing applications (Figure 6). In pixelRL, each pixel has an agent, and the agent changes the pixel value by taking an action. We also proposed an effective learning method for pixelRL that significantly improves the performance by considering not only the future states of the own pixel but also those of the neighbor pixels. The proposed method can be applied to some image processing tasks that require pixel-wise manipulations, where deep RL has never been applied. Besides, it is possible to visualize what kind of operation is employed for each pixel at each iteration, which would help us understand why and how such an operation is chosen. We applied the proposed method to a variety of image processing tasks: image denoising, image restoration, local color enhancement, and saliency-driven image editing. Our experimental results demonstrate that the proposed method achieves comparable or better performance, compared with the state-of-the-art methods based on supervised learning (Figure 6).

4 Realistic Augmented Reality

We presented RGB2AO [13], a novel task to generate ambient occlusion (AO) from a single RGB image instead of screen space buffers such as depth and normal. RGB2AO produces a new image filter that creates a non-directional shading effect that darkens enclosed and sheltered areas. RGB2AO aims to enhance two 2D image editing applications: image composition and geometry-aware contrast enhancement. We first collected a synthetic dataset consisting of pairs of RGB images and AO maps. Subsequently, we proposed a model for RGB2AO by supervised learning of a convolutional neural network (CNN), considering 3D geometry of the input image. Experimental results quantitatively and qualitatively demonstrate the effectiveness of our model (Figure 7). By using this technology, we can achieve more realistic augmented reality images for deeper immersion.

5 Applications

Though the works introduced in this paper are mostly on low-level image processing but have a lot of business applications. One of our most successful cases is the click rate prediction and improvement support for online banner images [14]. By showing the predicted scores and some design feedbacks, we can achieve better performing advertisement. In addition, our techniques can be applied to social media marketing [7]. For instance, we could increase the number of likes of a certain fast food brand account.

6 Conclusions

In this paper, we introduced some representative works of our attractiveness computing especially focusing on low-level image processing. We believe analyzing and enhancing the attractiveness of multimedia content is a promising direction with many real-world business applications.

References

[1] Toshihiko Yamasaki, Yusuke Fukushima, Ryosuke Furuta, L. Sun, Kiyoharu Aizawa, and D. Bollegala, "Prediction of User Ratings of Oral Presentations using Label Relations," ACMMM ASM, pp. 33-38, 2015.

[2] Toshihiko Yamasaki, Yusuke Fukushima, Ryosuke Furuta, and Kiyoharu Aizawa, "Towards Online Impression Prediction of Oral Presentations using Soft Coding," Proc. 1st workshop on Attractiveness Computing in Multimedia (ACM), pp. 462-465, 2016.

[3] Shengzhou Yi, Xueting Wang and Toshihiko Yamasaki, "Impression Prediction of Oral Presentation Using LSTM

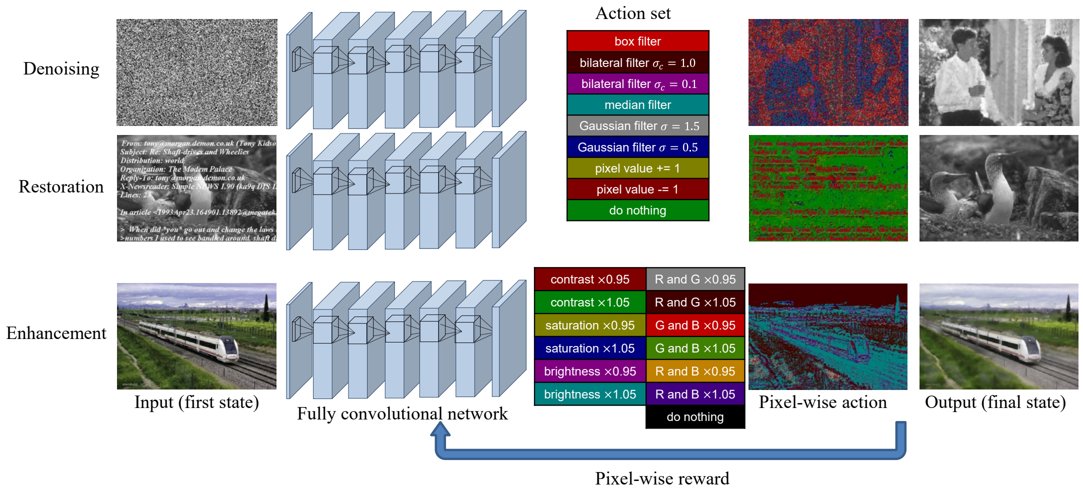

Figure 6 Overview of our method. In our framework, the generator is trained with RL to control image editing software, and the output of the discriminator is used as the reward.

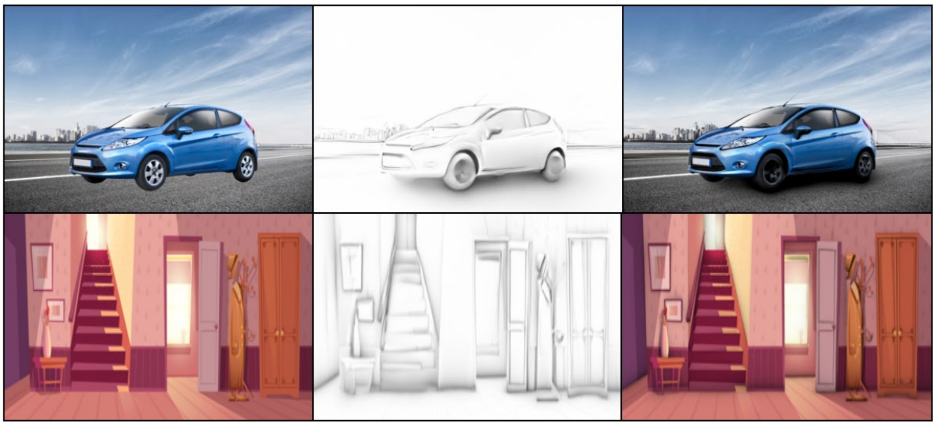

Figure 7 Ambient Occlusion (AO) generation from a single image with our RGB2AO framework. Given an RGB image (left), our model automatically generates an ambient occlusion map (center). We show two applications of applying the generated AO map effect: image composition (in the first row) and geometry-aware contrast enhancement (in the second row) on the right. The input images are from Adobe Stock.

and Dot-product Attention Mechanism," BigMM, short, pp. 242 – 246, 2019.

[4] Shinji Oyama and Toshihiko Yamasaki, "Visual Clarity Analysis and Improvement Support for Presentation Slides" BigMM, long, pp. 421-428, 2019. (invited)

[5] Toshihiko Yamasaki, Jiani Hu, Shumpei Sano, and Kiyoharu Aizawa, "FolkPopularityRank: Predicting and Enhancing Social Popularity using Text Tags in Social Networks," IJCAI, pp. 3231-3237, 2017.

[6] Xueting Wang, Yiwei Zhang, and Toshihiko Yamasaki, "User-Aware Folk Popularity Rank: User-Popularity-Based Tag Recommendation That Can Enhance Social Popularity," ACMMM, P4C-03, 2019.

[7] Yiwei Zhang and Toshihiko Yamasaki, "Style-Aware Image Recommendation for Social Media Marketing," ACMMM, 2021.

[8] Nobukatsu Kajiura, Satoshi Kosugi, Xueting Wang, and Toshihiko Yamasaki, "Self-Play Reinforcement Learning for Fast Image Retargeting," ACMMM, pp. 1755-1763, 2020.

[9] Ryosuke Furuta, Ikuko Tsubaki, and Toshihiko Yamasaki, "Fast Volume Seam Carving with Multi-pass Dynamic Programming," IEEE TCSVT, Volume 28, Issue 5, pp. 1087-1101, May 2018.

[10] Satoshi Kosugi and Toshihiko Yamasaki, "Unpaired Image Enhancement Featuring Reinforcement-Learning-Controlled Image Editing Software," AAAI, volume 34, no. 7, pp. 11296-11303, 2020.

[11] Ryosuke Furuta, Naoto Inoue, and Toshihiko Yamasaki, "Fully Convolutional Network with Multi-Step Reinforcement Learning for Image Processing," AAAI), Vol. 33, No. 1, pp. 3598-3605, 2019

[12] Ryosuke Furuta, Naoto Inoue, and Toshihiko Yamasaki, "PixelRL: Fully Convolutional Network with Reinforcement Learning for Image Processing," IEEE TMM, vol. 22, no. 7, pp. 1704-1719, July 2020.

[13] Naoto Inoue, Daichi Ito, Yannick Hold-Geoffroy, Long Mai, Brian Price, and Toshihiko Yamasaki, "RGB2AO: Ambient Occlusion Generation from RGB Images," CGF, vol. 39, issue 2, pp. 451-462, 2020.

[14] Bohui Xia, Hiroyuki Seshime, Xueting Wang, and Toshihiko Yamasaki, "Click-Through Rate Prediction of Online Banners Featuring Multimodal Analysis," International Journal of Semantic Computing (IJSC), Vol. 14, No. 1, pp. 71–91, 2020.