A Lightweight Auto-Crop Based on Deep Reinforcement Learning

Kunxiang Liu[1]*, Shaoqiang Zhu[1]†, and Junsong Zhang[1]‡

Central China Normal University, Wuhan, Hubei, China
695261623@qq.com, 1163285243@qq.com, zhangjs@xmu.edu

Abstract

We study the problem of image cropping, which aims at improving the aesthetic quality of images by cutting gradually from the edges around the image and re-composition. Correct composition is the key to high quality images. Most previous cropping approaches generate a great number of candidates box from input image and select the most pleasant one as the final cropped image which is time-consuming and may be an issue where the best cropping box is not in candidate box. To address these issues and motivate by these challenges, we propose a real-time and lightweight framework based on deep reinforcement learning algorithm, name advantage actor critic(A2C), to achieve fast and automatic cropping. Specifically, the sequential action of cropping is automatically learned through a policy network which contains a MobileNetV2 model, and the average intersection-over-union(IOU) value is designed as a part of learning reward. The model are trained by synchronous policy gradient and we show that parallel actor-learners, have an efficient learning on image cropping. Evaluating on the Flickr Cropping Dataset(FCD) and the experimental results show that our method reach the state-of-the-art performance with fewer cropping steps and time compared with some previous automatic cropping tools.

1 Introduction

Image cropping is to remove the unwanted outer areas from a photographic or illustrated image. This process usually involves deleting some peripheral areas of the image to remove excess waste from the picture, improving its frame, changing the aspect ratio, or emphasizing or isolating the subject matter from its background. Generally image cropping is the first step on image editing which call two-time composition in photography. Non-photography enthusiasts may not know how to compose or crop a photo correctly. Automatic cropping methods can help them easily get better picture compositions without having to master complex photographic composition techniques. The main difficulty in automatic image cutting of computer is that the first is to quantify the quality or aesthetics of pictures is a long-standing problem of computer vision [26], the second is cropping images have strong subjectivity that even for the same picture, the viewer may cut out different areas due to aesthetic differences. However, making computers learn to automatically crop images is challenging. Recently, most previous methods for automatic image cropping include attention-based and aesthetics-based methods and many relevant researchers have come up with innovative automatic cutting methods [9, 11, 30, 3]. The main difference between the these methods is that some methods regard the computer image cropping process as a determining-adjusting [29, 28] problem, and (1) first finds out the significant or most pleasant area and (2) generates a series of candidate diagrams based on this

*Designed and implemented the method
†Article writing
‡Provided a lot of suggestions

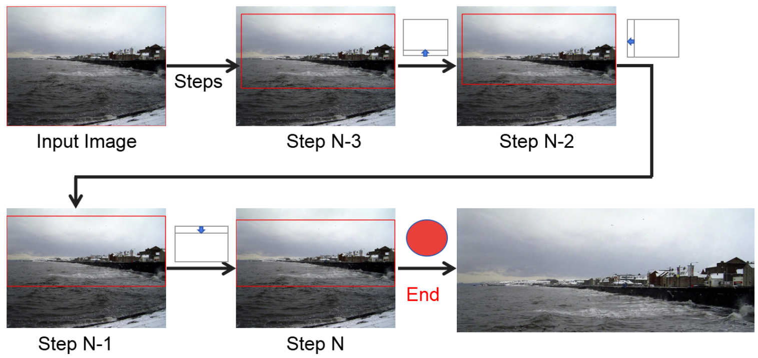

Figure 1: It shows the process of image cropping sequentially. Each step the agent take actions from current observation. And the environment performs cropping action on the image and the cropped image as the new observation. Repeat this process until the action is terminated or the number of cropped throttling is exceeded

using the sliding window method, and (3) selects the best perception image from the aesthetic evaluation model. However, the method is time consuming (the need to filter thousands of candidate images), the best cropping box is not the risk within the candidates. Some methods regard the image cropping process as the Markov decision-making process [12], simulating human to crop the image [16]. According to the global and local characteristics of the input picture, the model generates the corresponding cropping action, and the image is gradually cropped internally from the four-week edge until the output terminates the action or exceeds the limit (such as cutting up to 20 times, etc.). But aesthetic quality assessment or quantitative image aesthetics is a long-standing problem of computer vision. Lack of robustness of fractions as reward functions [26, 1]. Based on the above discussion, in this paper we propose a lightweight image cropping method, name LA2C, the method is based on deep enhanced learning algorithm Advantage Actor critic(A2C) [20]. We look the image clipping as the Markov decision-making process [12] and show the sequential cropping process in Figure 1. Our main contributions are summarized as follows:

1. Based on deep reinforcement learning, we propose an lightweight cropping Auto-Crop method which can fast and correct automatic image.

2. Abandoning the use of aesthetic score, which is difficult to accurately quantify the aesthetic quality of images as a reward. We use IOU value as part of the reward function.

3. Using the pre-trained MobileNetV2 [24] model to replace the common convolution layer for feature extraction, improve the ability to extract image features and accelerate training Simplify action space, including image clipping basic actions and a termination action.

2 Related work

Image cropping aims at improving the aesthetic quality by removing the unwanted outer areas from a photographic or illustrated image. Most previous cropping methods rely on the aesthetic quality assessment. We summarize representative works in image copping [29, 16, 28].

Recently, deep reinforcement learning has shown promising success in automatic image cropping. It [16] show that extracting high level features using CNNs and learning to crop photos with Asynchronous Advantage Actor-Critic algorithm can result in state-of-the-art well quality cropping performance. It performs that the automatic image cropping problem can be formulated as a sequential decision-making process and novel an Aesthetics Aware Reinforcement Learning(A2-RL) model for weakly supervised cropping problem. The model based on the asynchronous advantage actor-critic(A3C) algorithm. CNN layers extract high level features,and input 227×227 images, LSTM layer record the history observation and FC layers output the action. Then calculate the aesthetic scores as the part of reward function. But the key of this model is to find an appropriate metric to evaluate a photo precise aesthetic score. The traditional image evaluation metrics may not work well in this situation [31, 26, 1, 7, 15, 19]. Most previous methods for automatic image cropping include attention-based [3, 3, 21, 25] and aesthetics-based methods [10, 22]. Recently deep learning cropping framework combined attention and aesthetics components [29, 28, 17, 27], different from deep reinforcement learning, it formulate photo cropping as a determining-adjusting process. Attention model predict region locations where the most visually salient and generate 1,296 cropping candidates in total by using sling-window based on human attention map. Aesthetics-aware part select the highest aesthetically-score one as the final cropping. But to select the highest aesthetics value form 1,296 cropping candidates means each image needs to be calculated 1,296 times through the aesthetics model. Besides, the pleasing cropping window may not in these candidates generated based on visually salient map.

Early methods [6, 8, 14, 18] design handcrafted features relied on aesthetic knowledge. However, due to the greater subjectivity and diversity in the measurement of image aesthetics quality, it is difficult to determine the type and quantity of reliable features. Deep learning performs better on aesthetic assessment and image cropping [26, 1, 29, 16].

Deep reinforcement learning have been widely used in image caption [23], image editing [31] object detectio n [2, 13] etc. Photo cropping based on deep reinforcement learning was found result in state-of-the-art performance. We propose a novel system to achieve the auto-cropping of images within a DRL frame which performs better and faster.

3 Method

We formulate automatic image cropping as a sequential decision-making process and as agent-environment interaction problem, the Markov Decision Process(MDP) problem [12]. We propose a novel automatic cropping method based advantage actor Critic (A2C) algorithm [20]. Figure 2 shows the overall framework and process, the agent contains a policy network. The policy network generates a series of cropping actions based on the current input image, and samples the corresponding actions from the action area. Then the action space get the sampled action to interact with the environment. The shape of the image is cropped from the four-week edge. After the cropping action was executed in each step, the rolloutstorage stores the rewards returned by the environment for subsequent loss calculations. And the goal of the agent is to maximize the reward after each cropping. Next, the simple and lightweight framework will be described in detail from the environment, the agent, and the training process in three sections.

3.1 Environment

The role of the environment is as follows:

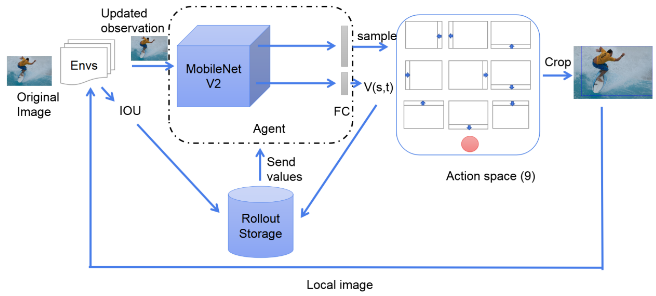

Figure 2: It shows the overall framework and process. The environment(env) generates a original image as a start observation and the mobilenetv2 [24] and the part of the the agent extract the features of cropped image from current observation. The features vector fed into the actor-critic part which contains two full-connection layers and output 9 actionn values and a state value. Then, the agent samples a cropping action from actions space and send it to the env. Furthermore, the env generates a reward(IOU value) based on the cropped image. The rolloutstorage stores the rewards returned by the environment for subsequent loss calculations

1. To provide the current observation O_0 for the agent. The completion of each cropping action will result in a change in the original image I_0, resulting in a new cropped image I_t, replacing the last observation $I_t \to O_t$. The advantage of using only cropped local images as observation instead of global futures to combine local futures is to reduce the number of duplicate pixel spaces and features and avoid wasting compute resources. Then, the input image resize to (224, 224) before entering the policy network.

2. Give a reward for performing a cropping action. The difference of cropping action will directly affect the difference of the next observation, and the reward of the corresponding action is given by the environment, which mimics the design of the Atari game environment. This is completely different from the previous deep intensive learning automatic cropping tool in the reward design, They use aesthetic quality assessment scores as rewards, but it is difficult to quantify accurately the aesthetic quality of a picture is a long-standing problem in computer vision. We propose use IOU value as the reward instead of aesthetic quality score cause IOU value correctly present the quality of cropping. However, the agent learn faster and more effective.

3. Action space and performing cropping actions. There are 9 actions in action space, 4 expansion actions, 4 zoom out actions, and 1 termination actions. Each action cropping stride is 1/30 high or wide for the image. The 1/30 stride can be cropped more accurately to the target box than the larger stride. The termination action means that the model will learn to decide when to terminate the cropping and will eventually crop the image output. Obviously, the cropping size is theoretically arbitrary.

In addition, the envs in the advantage actor critic algorithm is operated in parallel. The number of envs in this article is 16, and these envs run independently of each other and interact with the same agent. After running a certain number of step, our method synchronizes update across the network.

3.2 The Agent

The agent is the core part of the automatic cropping frame, in a nutshell, every step, the agent outputs an action according to the current observation, and passes the action to the envs, envs to crop the current image from the action space by selecting the corresponding cropping method. Below from the policy network, the loss function and the implementation details expand description.

3.3 Policy network

The policy network consists of a pre-trained mobilenetv2 [24] and two full-attached layers. Mobilenetv2 is a lightweight, efficient CNN model designed primarily for mobile device vision applications. It uses convolution that can be separated at depth as an efficient building block, Two new architectural features are introduced: 1) The linear bottleneck layer between layers, and 2) the connection shortcut between the bottleneck layers.

Drawing on the idea of migration learning, using CNN model ImageNet [7] pre-training as feature extraction module can effectively reduce training time and better training effect, and the comparison results will be shown in the experimental results. First, the current observation input is fed into the MobileNetv2 [24] image feature extraction model that removes the last layer, and the current feature graph is obtained, and the output-side parallel connection is passed to a FC with 9 nodes and a FC containing 1 nodes. The former outputs 9 action values, $output = [P(0), P(1), ..., P(8)]$, $p(t)$ indicates the possibility that action is T, which outputs the state value to evaluate the current observation expected reward $V(s_t)$.

3.3.1 The loss function

In order to get the best cropping effect, we give up the way that many of the previous method used aesthetic score as part of the reward function. The quality of quantitative image aesthetics is a difficult problem in computer vision for a long time. At present, the advanced quantitative model NIMA [26] not yet be able to accurately give the aesthetic score of each image. Therefore, in terms of stability and accuracy, we propose to use average Intersection-Over-Union (IOU) value as a cropped image that evaluates each step, the IOU value is naturally used as the reward. IOU values are the usual criteria for measuring the accuracy of cropping, and the specific calculation method will be explained in detail in implementation details. When the agent outputs a actions $a(s_t)$ based on the current observation, Env executes the action after getting a cropped image and calculates the corresponding IOU value as the reward R_t. R_t set as follows, $R(t) = +iou$ value when $\nabla r > 0$, $R(t) = -iou$ value when $r < 0$, $R(t) = 0$ otherwise. This means that each time you crop IOU value increases, the agent receives a reward and, conversely, a penalty that, when the output is terminated or exceeds the qualifying number of cropping steps, There is no reward. At this point, the reward for an image clipping process can be designed as Formulas (1):

$$r_t = \begin{cases} IOU & \nabla r > 0 \\ -IOU & \nabla r < 0 \\ 0 & \text{otherwise} \end{cases} \qquad \nabla = IOU_t - IOU_{t-1} \tag{1}$$

And the loss function is designed as Formulas(2)-(5):

$$loss = loss_{action} + \lambda loss_{value} - \beta loss_{dist} \tag{2}$$

$$loss_{action} = -\log p(a_t|s_t;\theta)(R_t - V(s_t;\theta_v)) \tag{3}$$

$$loss_{value} = \frac{\sum_{i=1}^{t}(R_i - V(s_i;\theta_v))^2}{t} \tag{4}$$

$$loss_{dist} = H(p(s_t;\theta)) \tag{5}$$

4 Experimental results

We first present the cropping process and the test database and then exhibit the results of this framework on the test set, using the same evaluation indicator average IOU value as the previous work [29, 16, 5] and average boundary displacement, in addition to increasing the average number of cropping steps per image and cutting time-consuming metrics.

4.1 CUHK-ICD

The CUHK-ICD [30] test set contains 150 images, Each image is a cropped window given by 3 photographers, respectively. The original images collect from Chinese University of Hong Kong's image cropping database [30]. When we get the final cropped image, we will calculate the IOU value and the BDE value with 3 groundtruth box respectively and record the statistics.

4.2 Flickr cropping dataset

The FCD [4] test set contains 374 images, Each image contains a manual callout box that calculates the resulting cropping window with the box to get the IOU value. Figure 3 shows the Deep-crop cropping results.

4.3 Evaluation metrics and results

To assess the capabilities of our methods(LA2C), test the IOU values, BDE values, cropping steps, cropping time, and other metrics on the CUHK-ICD [30] and FCD [4] test sets, as shown in table 1 and table 2.

The Boundary Displacement Error(BDE) is designed as the average displacement of four edges between the cropping box and the groundtruth rectangle:

$$BDE = \frac{\sum_i \|B_i^g - B_i^c\|}{4} \tag{6}$$

where $i \in \{left, right, bottom, up\}$ and $\{B_i\}_i$ means the edge of a groundtruth window or cropping window.The lower the BDE value, the better the cropping effect.

$$Avgstep = \sum_{i=1}^{n} \frac{step\_num_i}{n} \tag{7}$$

Where $step_num_i$represents the part i image clipping step, n represents the number of test images.

The cropping step is defined as the cropping step for each image from the start cropping to the end of the cumulative. The cropping step reflects the model cropping efficiency and

Figure 3: Example of sequence decision cropping

whether the optimal cropping order can be calculated. The fewer cropping steps, the higher the cropping efficiency.

Crop time is defined as the time it takes for each image to be cropped from start to finish, reflecting the speed at which the model is cropped. The shorter the cropping time, the faster the cropping speed.

Method	Annotation1		Annotation2		Annotation3	
	Avg loU	ADE	Avg loU	ADE	AloU	ADE
RankSVM+DecAF[4]	0.6643	0.0920	0.6556	0.0950	0.6439	0.0990
VFN+SW[5]	0.7401	0.0693	0.7187	0.0762	0.7132	0.0772
A2-RL[16]	0.8019	0.0524	0.7961	0.0535	0.7902	0.0535
LA2C-LSTM(Ours)	0.6873	0.0925	0.6992	0.0869	0.6846	0.0909
LA2C(Ours)	0.7850	0.0603	0.7806	0.0643	0.7758	0.0649

Table 1: Cropping Accuracy on CUHK-ICD [4]. The higher the AVG loU value, the lower the Avg Disp Error(ADE) value indicates the better the cropping result

4.4 Experiment Analysis

From the experimental data recorded in the table 1, it is obvious that our method(LA2C) performs great on the FCD database [4] test set, with the AVG IOU value, Avg disp error value, avg steps and AVG times four indicators fully ahead of the RankSVM-DECAFP [4], VFN++SW [5], A2-RL [16] method. From table 2, our method(LA2C) performs very close to A2-RL compared to the CUHK-ICD test set [30]. From table 3, our method(LA2C) performs

Method	Avg loU	Avg Disp
RankSVM+DeCAF[4]	0.6019	0.1060
VFN+SW++[5]	0.6442	0.0938
A2-RL[16]	0.6633	0.0892
LA2C-LSTM(Ours)	0.6714	0.0818
LA2C(Ours)	0.6844	0.0791

Table 2: Cropping Accuracy on FCD [4]. The higher the AVG IOU value, the lower the Avg disp error value indicates the better the cropping result

Method	Avg Steps	Avg Times(s)
VFN+SW++[5]	0.1125	5.446
A2-RL[16]	13.56	0.147
LA2C-LSTM(Ours)	9.29	0.067
LA2C(Ours)	7.96	0.027

Table 3: Cropping Accuracy on FCD [4]. The higher the AVG IOU value, the lower the Avg disp error value indicates the better the cropping result

better than the VFN+SW++ and A2-RL [16] model on the FCD [4] test set, especially the cutting step is reduced by an average of 5.6 times, the increase is 41.29%, the cropping time is shortened by 0.046s each image, and the increase is 31.29%.

Compared with the method based on the sliding window. These method selects out the most aesthetic image from a large number(1,125) of candidate windows. However, this cropping process is low efficiency and time consuming. Our method regret the cropping process as the Markov decision-making process and based on deep reinforcement learning take advantage of shorter cropping times and more anthropomorphic. Compared with the previous method based on deep reinforcement learning. Our method has a great advantage in cropping step and cropping time.

4.5 Limitations and future work

The proposed method suffers from a few limitations. One potential deficiency point is that our method(LA2C) does not combine professional photography with aesthetic knowledge, simply allowing the model to learn how to crop images on its own. This may result in cropped images with a large difference between cropping results and human aesthetics. On the other hand, the training samples in the database are positive samples and the model lacks negative sample learning. In the future, we will continue to study the problem of automatic image clipping and think about it from the following aspects. Make computer automatic cropping model more integrated into professional photography and aesthetic knowledge, and how to quantify and evaluate the aesthetic quality of images in model learning is helpful to further improve the cropping effect. In addition, we will try to migrate the method of automatic image cropping to video auto-cropping or composition.

5 Conclusion

In this paper, we regret the automatic image cropping problem as the Markov decision-making process [12] and propose a novel simple and lightweight method based on deep reinforcement learning algorithm, name Advantage Actor Critic(A2C) [20]. With the currently to accurate measurement of cropping effect indicators, IOU value reward and a network with strong ability to extract features, our LA2C method improve cropping accuracy and achieve real-time cropping while increasing the average step.

References

[1] Larbi Abdenebaoui, Benjamin Meyer, Albert Bruns, and Susanne Boll. Unna: A unified neural network for aesthetic assessment. In *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6. IEEE, 2018.

[2] Juan C Caicedo and Svetlana Lazebnik. Active object localization with deep reinforcement learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2488–2496, 2015.

[3] Jiansheng Chen, Gaocheng Bai, Shaoheng Liang, and Zhengqin Li. Automatic image cropping: A computational complexity study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 507–515, 2016.

[4] Yi-Ling Chen, Tzu-Wei Huang, Kai-Han Chang, Yu-Chen Tsai, Hwann-Tzong Chen, and Bing-Yu Chen. Quantitative analysis of automatic image cropping algorithms: A dataset and comparative study. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 226–234. IEEE, 2017.

[5] Yi-Ling Chen, Jan Klopp, Min Sun, Shao-Yi Chien, and Kwan-Liu Ma. Learning to compose with professional photographs on the web. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 37–45. ACM, 2017.

[6] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *European conference on computer vision*, pages 288–301. Springer, 2006.

[7] Yubin Deng, Chen Change Loy, and Xiaoou Tang. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine*, 34(4):80–106, 2017.

[8] Sagnik Dhar, Vicente Ordonez, and Tamara L Berg. High level describable attributes for predicting aesthetics and interestingness. In *CVPR 2011*, pages 1657–1664. IEEE, 2011.

[9] Seyed A Esmaeili, Bharat Singh, and Larry S Davis. Fast-at: Fast automatic thumbnail generation using deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4622–4630, 2017.

[10] Chen Fang, Zhe Lin, Radomir Mech, and Xiaohui Shen. Automatic image cropping using visual composition, boundary simplicity and content preservation models. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1105–1108. ACM, 2014.

[11] Eunbin Hong, Junho Jeon, and Seungyong Lee. Cnn based repeated cropping for photo composition enhancement. In *CVPR workshop*, 2017.

[12] Ronald A Howard. Dynamic programming and markov processes. 1960.

[13] Zequn Jie, Xiaodan Liang, Jiashi Feng, Xiaojie Jin, Wen Lu, and Shuicheng Yan. Tree-structured reinforcement learning for sequential object localization. In *Advances in Neural Information Processing Systems*, pages 127–135, 2016.

[14] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426. IEEE, 2006.

[15] Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. Photo aesthetics ranking network with attributes and content adaptation. In *European Conference on Computer Vision*, pages 662–679. Springer, 2016.

[16] Debang Li, Huikai Wu, Junge Zhang, and Kaiqi Huang. A2-rl: aesthetics aware reinforcement learning for image cropping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8193–8201, 2018.

[17] Debang Li, Junge Zhang, Kaiqi Huang, and Ming-Hsuan Yang. Composing good shots by exploiting mutual relations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4213–4222, 2020.

[18] Wei Luo, Xiaogang Wang, and Xiaoou Tang. Content-based photo quality assessment. In *2011 International Conference on Computer Vision*, pages 2206–2213. IEEE, 2011.

[19] Long Mai, Hailin Jin, and Feng Liu. Composition-preserving deep photo aesthetics assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 497–506, 2016.

[20] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pages 1928–1937, 2016.

[21] Naila Murray, Luca Marchesotti, and Florent Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2408–2415. IEEE, 2012.

[22] Masashi Nishiyama, Takahiro Okabe, Yoichi Sato, and Imari Sato. Sensation-based photo cropping. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 669–672. ACM, 2009.

[23] Zhou Ren, Xiaoyu Wang, Ning Zhang, Xutao Lv, and Li-Jia Li. Deep reinforcement learning-based image captioning with embedding reward. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 290–298, 2017.

[24] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[25] Bongwon Suh, Haibin Ling, Benjamin B Bederson, and David W Jacobs. Automatic thumbnail cropping and its effectiveness. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 95–104. ACM, 2003.

[26] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8):3998–4011, 2018.

[27] Yi Tu, Li Niu, Weijie Zhao, Dawei Cheng, and Liqing Zhang. Image cropping with composition and saliency aware aesthetic score map. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12104–12111, 2020.

[28] Wenguan Wang and Jianbing Shen. Deep cropping via attention box prediction and aesthetics assessment. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2186–2194, 2017.

[29] Wenguan Wang, Jianbing Shen, and Haibin Ling. A deep network solution for attention and aesthetics aware photo cropping. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[30] Jianzhou Yan, Stephen Lin, Sing Bing Kang, and Xiaoou Tang. Learning the change for automatic image cropping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 971–978, 2013.

[31] Runsheng Yu, Wenyu Liu, Yasen Zhang, Zhi Qu, Deli Zhao, and Bo Zhang. Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning. In *Advances in Neural Information Processing Systems*, pages 2153–2163, 2018.