

# A2-RL: Aesthetics Aware Reinforcement Learning for Image Cropping

Debang Li<sup>1,2</sup>, Huikai Wu<sup>1,2</sup>, Junge Zhang<sup>1,2</sup>, Kaiqi Huang<sup>1,2,3</sup>

<sup>1</sup> CRIPAC & NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup> CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing, China

{debang.li, huikai.wu, jgzhang, kaiqi.huang}@nlpr.ia.ac.cn

## Abstract

Image cropping aims at improving the aesthetic quality of images by adjusting their composition. Most weakly supervised cropping methods (without bounding box supervision) rely on the sliding window mechanism. The sliding window mechanism requires fixed aspect ratios and limits the cropping region with arbitrary size. Moreover, the sliding window method usually produces tens of thousands of windows on the input image which is very time-consuming. Motivated by these challenges, we firstly formulate the aesthetic image cropping as a sequential decision-making process and propose a weakly supervised Aesthetics Aware Reinforcement Learning (A2-RL) framework to address this problem. Particularly, the proposed method develops an aesthetics aware reward function which especially benefits image cropping. Similar to human’s decision making, we use a comprehensive state representation including both the current observation and the historical experience. We train the agent using the actor-critic architecture in an end-to-end manner. The agent is evaluated on several popular unseen cropping datasets. Experiment results show that our method achieves the state-of-the-art performance with much fewer candidate windows and much less time compared with previous weakly supervised methods.

## 1. Introduction

Image cropping is a common task in image editing, which aims to extract well-composed regions from ill-composed images. It can improve the visual quality of images, because the composition plays an important role in the image quality. An excellent automatic image cropping algorithm can give editors professional advices and help them save a lot of time [14].

In the past decades, many researchers have devoted their efforts to proposing novel methods [34, 10, 12] for automatic image cropping. As the cropping box annotations are expensive to obtain, several weakly supervised cropping

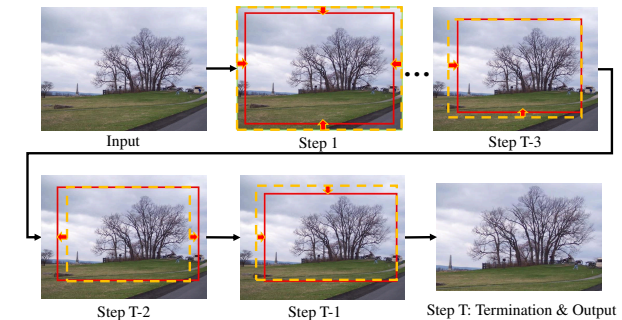


Figure 1. Illustration of the sequential decision-making based automatic cropping process. The cropping agent starts from the whole image and takes actions to find the best cropping window in the input image. At each step, it takes an action (yellow and red arrow) and transforms the previous window (dashed-line yellow rectangle) to a new state (red rectangle). The agent takes the termination action and stops the cropping process to output the cropped image at step T.

methods (without bounding box supervision) [11, 5, 35] are proposed. Most of these weakly supervised methods follow a three-step pipeline: 1) Densely extract candidates with the sliding window method on the input image, 2) Extract carefully designed features from each region and 3) Use a classifier or ranker to grade each window and find the best region. Although these works have achieved pretty good performance, they may not find the best results due to the limitations of the sliding window method, which requires fixed aspect ratios and limits the cropping region with arbitrary size. What’s more, these sliding window based methods usually need tens of thousands of candidates on image level, which is very time-consuming. Although we can set several different aspect ratios and densely extract candidates, it inevitably costs lots of time and is still unable to cover all conditions.

Based on above observations, in this paper, we firstly formulate the automatic image cropping problem as a sequential decision-making process, and propose an Aesthetics

Aware Reinforcement Learning (A2-RL) model for weakly supervised cropping problem. The sequential decision-making based automatic image cropping process is illustrated in Figure 1. To our knowledge, we are the first to put forward a reinforcement learning based method for automatic image cropping. The A2-RL model can finish the cropping process within several or a dozen steps and get results of almost arbitrary shape, which can overcome the disadvantages of the sliding window method. Particularly, A2-RL model develops a novel aesthetics aware reward function which especially benefits image cropping. Inspired by human’s decision making, the historical experience is also explored in the state representation to assist the current decision. We test the model on three unseen popular cropping datasets [34, 11, 4], and the experiment results demonstrate that our method obtains the state-of-the-art cropping performance with much fewer candidate windows and much less time compared with related methods.

## 2. Related Work

Image cropping aims at improving the composition of images, which is very important for the aesthetic quality. There are a number of previous works for aesthetic quality assessment. Many early works [15, 7, 19, 9] focus on designing handcrafted features based on intuitions from human’s perception or photographic rules. Recently, thanks to the fast development of deep learning and newly proposed large scale datasets [22], there are many new works [16, 20, 8] which accomplish aesthetic quality assessment with convolutional neural networks.

Previous automatic image cropping methods can be divided into two classes, attention-based and aesthetics-based methods. The basic approach of attention-based methods [28, 27, 24, 2] is to find the most visually salient regions in the original images. Attention-based methods can find cropping windows that draw more attention from people, but they may not generate very pleasing cropping windows, because they hardly consider about the image composition [4]. For those aesthetics-based methods, they aim to find the most pleasing cropping windows from original images. Some of these works [23, 11] use aesthetic quality classifiers to discriminate the quality of candidate windows. Other works use RankSVM [4] or RankNet [5] to grade each candidate window. There are also change-based methods [34], which compares original images with cropped images so as to throw away distracting regions and retain high quality ones. Image retargeting techniques [6, 3] adjust the aspect ratio of an image to fit the target aspect ratio, while not discarding important content in an image, which are relevant to our task.

As for the supervision information, these methods can be divided into supervised and weakly supervised methods, depending on whether they use bounding box annotations.

Supervised cropping methods [12, 10, 31, 32] need bounding box annotations to train the cropper. For example, object detection based cropping methods [10, 32] are fast and effective, but they need a mount of bounding box annotations for training the detector, which is expensive. Most weakly supervised methods [11, 5, 14] still rely on the sliding window method to obtain the candidate windows. As discussed above, the sliding window method uses fixed aspect ratios and limits windows with arbitrary size. What’s more, these methods are also very time-consuming. In this paper, we formulate the cropping process as a sequential decision-making process and propose a weakly supervised reinforcement learning (RL) based strategy to search the cropping window. Hong *et al.* [12] also regard the cropping process as a sequential process, but they use bounding box as supervision. Our RL based method can find the final results with only several or a dozen candidates of almost arbitrary size, which is much faster and more effective compared to other weakly supervised methods and doesn’t need bounding box annotations compared to supervised methods.

RL based strategies have been successfully applied in many domains of computer vision, including image caption [26], object detection [1, 13] and visual relationship detection [18]. The active object localization method [1] achieves the best performance among detection algorithms without region proposals. The tree-RL method [13] uses RL to obtain region proposals and achieves comparable result with much fewer region proposals compared to RPN [25]. Above RL based object detection methods use bounding boxes as their supervision, however, our framework only uses the aesthetics information as supervision, which requires less label information. To our best knowledge, we are the first to put forward a deep reinforcement learning based method for automatic image cropping.

## 3. Aesthetics Aware Reinforcement Learning

In this paper, we formulate automatic image cropping as a sequential decision-making process. In the decision-making process, an agent interacts with the environment, and takes a series of actions to optimize a target. As illustrated in Figure 2, for our problem, the agent receives observations from the input image and the cropping window. Then it samples action from the action space according to the observation and historical experience. The agent executes the sampled action to manipulate the shape and position of the cropping window. After each action, the agent receives a reward according to the aesthetic score of the cropped image. The agent aims to find the most pleasing window in the original image by maximizing the accumulated reward. In this section, we first introduce the state space, action space and aesthetics aware reward of our model, then we detail the architecture of our aesthetics aware reinforcement learning (A2-RL) model and the

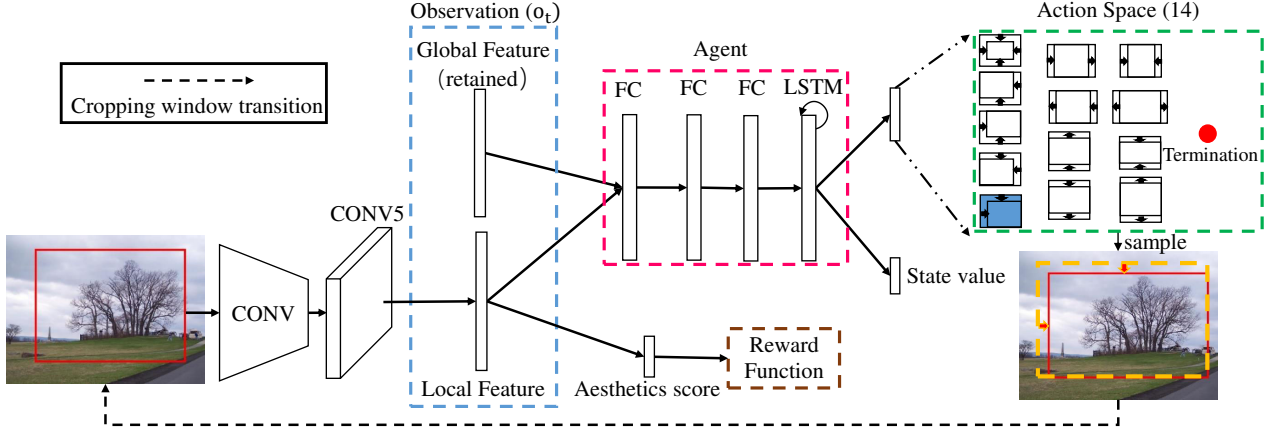


Figure 2. Illustration of the A2-RL model architecture. In the forward pass, the feature of the cropping window (local feature) is extracted and concatenated with the feature of the whole image (global feature) which is extracted and retained previously. Then, the concatenated feature vector is fed into the actor-critic branch which has two outputs. The actor output is used to sample actions from the action space so as to manipulate the cropping window. The critic output (state value) is used to estimate the expected reward under the current state. In addition, the feature of the cropping window is also fed into the aesthetic quality assessment branch. The output of this branch is the aesthetic score of the input cropping window and stored to compute rewards for actions. In this model, both the global feature and the local feature are 1000-dim vectors, three fully-connected layers and the LSTM layer all output 1024-dim feature vectors.

training process.

### 3.1. State and Action Space

At each step, the agent decides which action to execute according to the current state. The state must provide the agent with comprehensive information for better decisions. As the A2-RL model formulates the automatic image cropping as a sequential decision-making process, the current state can be represented as  $s_t = \{o_0, o_1, \dots, o_{t-1}, o_t\}$ , where  $o_t$  is the current observation of the agent. This formulation is similar to human’s decision making process, which considers not only the current observation but also the historical experience. The historical experience is usually very valuable for future decision-making. Thus, in the proposed method, we also take the historical experience into consideration. The A2-RL model uses the features of the cropping window and the input image as the current observation  $o_t$ . Agent can learn about the global information and the local information with such observation. In the A2-RL model, we use a LSTM unit to memorize historical observations  $\{o_0, o_1, \dots, o_{t-1}\}$ , and combine them with the current observation  $o_t$  to form the state  $s_t$ .

We choose 14 pre-defined actions to form the action space, which can be divided into four groups: scaling actions, position translation actions, aspect ratio translation actions and a termination action. The first three groups aim to adjust the size, position and shape of the cropping window, including 5, 4 and 4 actions respectively. These three groups follow similar definitions in [13], but with different scales. All these actions adjust the shape and position by

0.05 times of the original image size, which could capture more accurate cropping windows than a large scale. The termination action is a trigger for the agent, when this action is chosen, the agent will stop the cropping process and output the current cropping window as the final result. As the model learns when to stop the cropping process by itself, it can stop at the state where the score won’t increase anymore so as to get the best cropping window. Theoretically, the agent can cover windows with almost arbitrary size and position on the original image.

The observation and action space are illustrated in Figure 2 for an intuitional representation.

### 3.2. Aesthetics Aware Reward

Our A2-RL model aims to find the most pleasing cropping window on the original image. So the reward function should lead the agent to find a more pleasing window at each step. We propose using the aesthetic score to evaluate the pleasing degree of images naturally. When the agent takes an action, the difference between the aesthetic scores of the new cropping window and the last one can be utilized to compute the reward for this action. More detailed, if the aesthetic score of the new window is higher than the last one, the agent will get a positive reward. On the contrary, if the score becomes lower, the agent will get a negative reward. To speed up the cropping process, we also give the agent an additional negative reward  $-0.001 * (t + 1)$  at each step, where  $t + 1$  is the number of steps the agent has taken since the beginning and  $t$  starts from 0. This constraint will result in a lower reward when the agent takes too

many steps. For an image  $I$ , we denote its aesthetic score as  $s_{aes}(I)$ . The new cropped image and the last one are denoted as  $I_{t+1}$  and  $I_t$  respectively,  $sign(*)$  denote the sign function, so the foundation of our aesthetics aware reward function  $r'_t$  can be formulated as :

$$r'_t = sign(s_{aes}(I_{t+1}) - s_{aes}(I_t)) - 0.001 * (t + 1) \quad (1)$$

In the above definition of  $r'_t$ , we use the sign function to limit the variation range of  $s_{aes}(I_{t+1}) - s_{aes}(I_t)$ , because the training is stable and easy to converge in practice under such setting. Using the reward function without the sign function makes it hard for the model to converge in our experiments, which is mainly due to the dramatic fluctuation of rewards, especially when the model samples the cropping window randomly at first.

We also consider other heuristic constraints for better cropping policies. We believe the aspect ratio of well-composed images is limited in a particular range. In the A2-RL model, if the aspect ratio of the new window is lower than 0.5 or higher than 2, the agent will receive a negative reward  $nr$  as the penalty term for the corresponding action, so the agent can learn a strict rule not to let such situation happen. The limited range of the aspect ratio in our model is for the common cropping task, we can also modify the reward function and the action space to meet some special requirements on the aspect ratio depending on the application. Let  $ar$  denote the aspect ratio of the new window,  $nr$  denote the negative reward the agent receives when the aspect ratio of the window exceeds the limited range, the whole reward function  $r_t$  for the agent taking an action  $a_t$  under the state  $s_t$  can be formulated as:

$$r_t(s_t, a_t) = \begin{cases} r'_t + nr, & \text{if } ar < 0.5 \text{ or } ar > 2 \\ r'_t, & \text{otherwise} \end{cases} \quad (2)$$

### 3.3. A2-RL Model

With the defined state space, action space and reward function, we start to introduce the architecture of our Aesthetics Aware Reinforcement Learning (A2-RL) framework. The detailed architecture of the framework is illustrated in Figure 2. The A2-RL model starts with a 5-layer convolution block and a fully-connected layer which outputs 1000-dimensional vector for feature representation. Then the model splits into two branches, the first one is the actor-critic branch, the other is the aesthetic quality assessment branch. The actor-critic branch is composed of three fully-connected layers and a LSTM layer. The LSTM layer is used to memorize the historical observations. The actor-critic branch has two outputs, the first one is the policy output, which is also named **Actor**, the other output is the value output, also named **Critic**. The policy output is a fourteen-dimensional vector, each dimension corresponding to the probability of taking relevant action. The value output is

the estimation of the current state, which is the expected accumulated reward in the current situation. The aesthetic quality assessment branch outputs an aesthetic quality score for the cropped image, which is used to compute the reward.

In the image cropping process, the A2-RL model provides the agent with the probability of each action under the current state. As shown in Figure 2, the model feeds the cropped image into the feature representation unit and extracts the local feature at first. Then the feature is combined with the global feature which is extracted in the first forward pass and retained for the following process. The combined feature vector is then fed into the actor-critic branch. According to the policy output, the agent samples the relevant action and adjusts the size and position of the cropping window correspondingly. For example, in Figure 2, the agent executes the sampled action to shrink the cropping window from left and top with 0.05 times the size of the image. Forward pass will continue until the termination action is sampled.

### 3.4. Training A2-RL Model

In the A2-RL, we propose using the asynchronous advantage actor-critic (A3C) algorithm [21] to train the cropping policy. Different from the original A3C, we replace the asynchronous mechanism with mini-batch to increase the diversity. In the training stage, we use the advantage function [21] and entropy regularization term [33] to form the optimization objective of the policy output. We use  $R_t$  to denote the accumulated reward at step  $t$ , which is  $\sum_{i=0}^{k-1} \gamma^i r_{t+i} + \gamma^k V(s_{t+k}; \theta_v)$ , where  $\gamma$  is the discount factor,  $r_t$  is the aesthetics aware reward at step  $t$ ,  $V(s_t; \theta_v)$  is the value output under state  $s_t$ ,  $\theta_v$  denotes the network parameters of **Critic** branch and  $k$  ranges from 0 to  $t_{max}$ .  $t_{max}$  is the maximum number of steps before updating. The optimization objective of the policy output is to maximize the advantage function  $R_t - V(s_t; \theta_v)$  and the entropy of the policy output  $H(\pi(s_t; \theta))$ , where  $\pi(s_t; \theta)$  is the probability distribution of policy output,  $\theta$  denotes the network parameters of **Actor** branch, and  $H(*)$  is the entropy function. The entropy in the optimization objective is used to increase the diversity of actions, which can make the agent learn flexible policies. The optimization objective of the value output is to minimize  $(R_t - V(s_t; \theta_v))^2/2$ . So gradients of the actor-critic branch can be formulated as  $\nabla_{\theta} \log \pi(a_t | s_t; \theta) (R_t - V(s_t; \theta_v)) + \beta \nabla_{\theta} H(\pi(s_t; \theta))$  and  $\nabla_{\theta_v} (R_t - V(s_t; \theta_v))^2/2$ , where  $\beta$  is to control the influence of entropy and  $\pi(a_t | s_t; \theta)$  is the probability of the sampled action  $a_t$  under the state  $s_t$ .

The whole training procedure of the A2-RL model is described in Algorithm 1.  $T_{max}$  means maximum number of steps the agent takes before termination.

**Algorithm 1:** Training procedure of the A2-RL model

---

**Input:** original image  $I$

- 1  $f_{global} = Feature\_extractor(I)$
- 2  $I_0 \leftarrow I, t \leftarrow 0$
- 3 **repeat**
- 4    $t_{start} = t, d\theta \leftarrow 0, d\theta_v \leftarrow 0$
- 5   **repeat**
- 6      $f_{local} = Feature\_extractor(I_t)$
- 7      $o_t = concat(f_{global}, f_{local})$
- 8      $s_t = LSTM_{AC}(o_t)$  //LSTM of Actor-Critic
- 9     Perform  $a_t$  according to the policy output
- 10     $\pi(a_t|s_t; \theta)$  and get the new image  $I_{t+1}$
- 11     $r_t = reward(I_t, I_{t+1}, t)$
- 12     $t = t + 1$
- 13    **until**  $t - t_{start} == t_{max}$  or  $a_{t-1}$  is termination action;
- 14     $R = \begin{cases} 0 & \text{if } a_{t-1} \text{ is termination action} \\ V(s_t; \theta_v) & \text{for other actions} \end{cases}$
- 15    **for**  $i \in \{t-1, \dots, t_{start}\}$  **do**
- 16      $R \leftarrow r_i + \gamma R$
- 17      $d\theta \leftarrow d\theta + \nabla_{\theta} \log \pi(a_i|s_i; \theta)(R - V(s_i; \theta_v)) + \beta \nabla_{\theta} H(\pi(s_i; \theta))$
- 18      $d\theta_v \leftarrow d\theta_v + \nabla_{\theta_v} (R - V(s_i; \theta_v))^2 / 2$
- 19    **end**
- 20    Update  $\theta$  with  $d\theta$  and  $\theta_v$  with  $d\theta_v$
- 21 **until**  $t == T_{max}$  or  $a_{t-1}$  is termination action;

---

## 4. Experiments

### 4.1. Experimental Settings

**Training Data** To train our network, we select images from a large scale aesthetics image dataset named AVA [22], which consists of  $\sim 250000$  images. All these images are labeled with aesthetic score rating from one to ten by several people. As the score distribution is extremely unbalanced, we simply divide them into three classes: low quality, middle quality and high quality. These three classes correspond to score from one to four, four to seven and seven to ten respectively. We choose about 3000 images from each class to compose the training set. Finally, there are  $\sim 9000$  images in the training set. With these training data, our model can learn policies with images of diverse quality, which can make the model generalize well to different images.

**Implementation Details** In our experiment, the aesthetic score  $s_{aes}(I)$  of the image  $I$  is the output of the pre-trained view finding network (VFN) [5], which is an aesthetic ranker modified from the original AlexNet [17]. The VFN is trained with the same training data and ranking loss as the original settings [5]. As shown in Figure 2, the actor-critic branch share the feature extractor unit with the VFN.

Method	Avg IoU	Avg Disp Error
eDN [30]	0.4857	0.1372
RankSVM+DeCAF <sub>7</sub> [4]	0.6019	0.1060
VFN+SW [5]	0.6328	0.0982
A2-RL w/o $nr$	0.5720	0.1178
A2-RL w/o LSTM	0.6310	0.1014
A2-RL(Ours)	<b>0.6633</b>	<b>0.0892</b>

Table 1. Cropping Accuracy on FCD [4].

RMSProp [29] algorithm is utilized to optimize the A2-RL model, the learning rate is set to 0.0005 and the other arguments are set by default values. The mini batch size in training is set to 32. The discount factor  $\gamma$  is set as 0.99 and the weight of entropy loss  $\beta$  is set as 0.05 respectively. The  $T_{max}$  is set as 50, and the update period  $t_{max}$  is set to 10. The penalty term  $nr$  in reward function is empirically set to -5, which can lead to a strict rule that prevents the aspect ratio of the cropping window exceeding the limited range.

We also choose 900 images from AVA dataset as the validation set following the way of the training set. As the A2-RL model aims to find the cropping window with the highest aesthetic score, on the validation set, we use the improvement of aesthetic score between the original and cropped images as metric. We train the networks on the training set for 20 epochs and validate the models on the validation set every epoch. The model which achieves the best average improvement on the validation set is chosen as the final model.

**Evaluation Data and Metrics** To evaluate the capacities of our agent, we test it on three unseen automatic image cropping datasets, including CUHK Image Cropping Dataset (CUHK-ICD) [34], Flickr Cropping Dataset (FCD) [4] and Human Cropping Dataset (HCD) [11]. The first two datasets use the same evaluation metrics, while the last one uses different metrics. We adopt the same metrics as the original works for fair comparison.

There are 950 test images in CUHK-ICD, which are annotated by three different expert photographers. FCD contains 348 test images, and each image has only one annotation. On these two datasets, previous works [34, 4, 5] use the same evaluation metrics to measure the cropping accuracy, including *average intersection-over-union (IoU)* and *average boundary displacement*. In this paper, we denote the ground truth window of the image  $i$  as  $W_i^g$  and the cropping window as  $W_i^c$ . The average *IoU* of  $N$  images can be computed as

$$1/N \sum_{i=1}^N area(W_i^g \cap W_i^c) / area(W_i^g \cup W_i^c) \quad (3)$$

The average boundary displacement computes the average distance between the four edges of the ground truth win-

Method	Annotation I		Annotation II		Annotation III	
	Avg IoU	Avg Disp Error	Avg IoU	Avg Disp Error	Avg IoU	Avg Disp Error
eDN [30]	0.4636	0.1578	0.4399	0.1651	0.4370	0.1659
RankSVM+DeCAF <sub>7</sub> [4]	0.6643	0.092	0.6556	0.095	0.6439	0.099
LearnChange [34]	0.7487	0.0667	0.7288	0.0720	0.7322	0.0719
VFN+SW [5]	0.7401	0.0693	0.7187	0.0762	0.7132	0.0772
A2-RL w/o <i>nr</i>	0.6841	0.0852	0.6733	0.0895	0.6687	0.0895
A2-RL w/o LSTM	0.7855	0.0569	0.7847	0.0578	0.7711	0.0578
A2-RL(Ours)	<b>0.8019</b>	<b>0.0524</b>	<b>0.7961</b>	<b>0.0535</b>	<b>0.7902</b>	<b>0.0535</b>

Table 2. Cropping Accuracy on CUHK-ICD [34].

dow and the cropping window. In image  $i$ , we denote four edges of the ground truth window as  $B_i^g(l)$ ,  $B_i^g(r)$ ,  $B_i^g(u)$ ,  $B_i^g(b)$ , correspondingly, four edges of the cropping window are denoted as  $B_i^c(l)$ ,  $B_i^c(r)$ ,  $B_i^c(u)$ ,  $B_i^c(b)$ . The *average boundary displacement* of  $N$  images can be computed as

$$1/N \sum_{i=1}^N \sum_{j=\{l,r,u,b\}} |B_i^g(j) - B_i^c(j)|/4 \quad (4)$$

HCD contains 500 test images, each is annotated by ten people. Because it has more annotations for each image than the first two datasets, the evaluation metric is a little different. Previous works [11, 14] on this dataset use *top-K maximum IoU* as the evaluation metric, which is similar to the previous average IoU. *Top-K maximum IoU* metric computes the IoU between the proposed cropping windows and ten ground truth windows, then it chooses the maximum IoU as the final result. *Top-k* means to use  $k$  best cropping windows to compute the result.

## 4.2. Evaluation of Cropping Accuracy

In this section, we compare the cropping accuracy of our A2-RL model with previous sliding window based weakly supervised methods to validate its effectiveness. As the aesthetic assessment of our model is based on VFN [5], we mainly compare our model with this method. Our model uses RL based method to search the best cropping windows sequentially with only several candidates. The VFN-based method uses sliding window to densely extract candidates. We also compare with several other baselines.

**Cropping Accuracy on CUHK-ICD and FCD** As the previous VFN method [5] is only evaluated on CUHK-ICD [34] and FCD [4], we also mainly compare our framework with VFN on these two datasets. Notably, the original VFN not only uses the sliding window candidates, but also uses the *ground truth* window of test images as candidates, which leads to a remarkably high performance on these two datasets. As A2-RL model aims to search the best cropping window, and in practice, there won't be any ground truth window for cropping algorithms, so, in this experiment, we

don't use any ground truth windows in both frameworks for fair comparison. It's also worthy to mention that, the A2-RL model has never seen images from both datasets during training.

Besides the two frameworks discussed above, we also compare some other cropping methods. We choose the best attention-based method eDN reported in [4] on behalf of the attention-based cropping algorithms. This method computes the saliency maps with algorithms from [30], and search the best cropping window by maximizing the difference of average saliency between the cropping window and other region. We also choose the best result (*RankSVM+DeCAF<sub>7</sub>*) reported in [4] as another baseline. In this method, aesthetic feature *DeCAF<sub>7</sub>* is extracted from AlexNet and a *RankSVM* is trained to find the best cropping window among all the candidates. For all these sliding window based methods, including *eDN*, *RankSVM+DeCAF<sub>7</sub>* and *VFN+SW (sliding window)*, the results are all reported with the same sliding window setting as [4].

Experiments on FCD are shown in Table 1, where *VFN+SW* and *A2-RL* are the two mainly comparable frameworks. We also show the results on CUHK-ICD in Table 2. As there are 3 annotations for each image, following previous works [34, 4, 5], we list the results for each annotation separately. All symbols in Table 2 are the same as Table 1. What's more, we also report the best result in [34], in which this dataset is proposed. Notably, the method is trained with supervised cropping data on this dataset, which is not very fair for us to compare. As this method is change-based, we denote it as *LearnChange* in Table 2.

From Tables 1 and 2, we can see that our A2-RL model outperforms other methods consistently on these two datasets, which demonstrates the effectiveness of our model.

**Cropping Accuracy on HCD** We also evaluate our A2-RL model on HCD [11]. Following previous works [11, 14] on this dataset, *top-K maximum IoU* is employed as the metric of cropping accuracy. We choose two state-of-the-art methods [11, 14] on this dataset as our baselines. The results are shown in Table 3. As our A2-RL model finds one

Method	Top-1 Max IoU
Fang <i>et al.</i> [11]	0.6998
Kao <i>et al.</i> [14]	0.7500
A2-RL w/o <i>nr</i>	0.7089
A2-RL w/o LSTM	0.7960
A2-RL(Ours)	<b>0.8204</b>

Table 3. Cropping Accuracy on HCD [11].

Method	Avg IoU	Avg Disp	Avg Steps	Avg Time(s)
VFN+SW	0.6328	0.0982	137	1.29
VFN+SW+	0.6395	0.0956	500	4.37
VFN+SW++	0.6442	0.0938	1125	9.74
A2-RL(Ours)	<b>0.6633</b>	<b>0.0892</b>	<b>13.56</b>	<b>0.245</b>

Table 4. Time Efficiency comparison on FCD [4]. VFN+SW, VFN+SW+ and VFN+SW++ correspond different number of candidate windows, where VFN+SW follows original setting [5].

cropping window at a time, we compare the results using the *top-1 Max IoU* as metric. From Table 3, we can see that our A2-RL model still outperforms the state-of-the-art methods.

### 4.3. Evaluation of Time Efficiency

In this section, we study the time efficiency of our A2-RL model. We compare our model with the sliding window based VFN model on FCD. Experimental results are shown in Table 4. The *Avg Steps* and *Avg Time* mean the average value of steps and time methods cost to finish the cropping process on a single image. We also augment the number of sliding windows in this experiment. Notably, all results in Table 4 are evaluated on the same machine, which has a single NVIDIA GeForce Titan X pascal GPU with 12GB memory and Intel Core i7-6800k CPU with 6 cores.

From Table 4, we can easily find that the cropping accuracy is improved as we augment the number of sliding windows, but the consumed time also grows. Unsurprisingly, our A2-RL model needs much fewer steps and costs much less time than other methods. The average number of steps our A2-RL model takes is more than 10 times less than the sliding window based methods, but our A2-RL model still gets better cropping accuracy. These results show the capacities of our RL-based model, with the novel aesthetics aware reward and history-preserved state representation, our model learns to use as few actions as possible to obtain a more pleasant image.

### 4.4. Experiment Analysis

In this section, we analyse the experiment results and study our model.

**RL Search vs. Sliding Window** From Tables 1, 2 and 4, we can find out that the A2-RL method is better than the

VFN+SW method in cropping accuracy and time efficiency consistently. The main difference between these two methods is the way to get the cropping candidates. From this observation, we conclude that our proposed RL-based search method is better than the sliding window method, which is very obvious. Although the sliding window method can densely extract candidates, it still fails to find very accurate candidates due to the fixed aspect ratios. On the contrary, our A2-RL model can find cropping windows with almost arbitrary size.

### Observation+History Experience vs. only Observation

We use LSTM unit to memorize historical observations  $\{o_0, o_1, \dots, o_{t-1}\}$  and combine them with the current observation  $o_t$  to form the state  $s_t$ . In this section, we study the effect of the history experience in our model. We abandon the LSTM unit in the A2-RL model, so the agent only uses the current observation  $o_t$  as the state  $s_t$  to make decisions. We train a new agent under such setting and evaluate it on above three datasets. Results are shown in Tables 1, 2 and 3, where the new agent is denoted as A2-RL w/o LSTM. From these results, we can find that the cropping accuracy of the new model is much lower than the original A2-RL model, which demonstrates the importance of historical experiences.

**The effect of the limited aspect ratio.** As shown in Equation 2, if the aspect ratio of the cropped image exceeds the limited range, the agent will get an additional negative reward *nr*. In this section, we study the effect of the penalty term *nr* in the reward function. We remove the penalty term *nr* in the reward function and train a new agent. The new agent is evaluated on the above three datasets and the results are shown in Tables 1, 2 and 3, where the new agent is denoted as A2-RL w/o *nr*. From these results, we can find that the cropping accuracy of the new agent also decreases a lot, which demonstrates the importance of the penalty term *nr* in the reward function.

### 4.5. Qualitative Analysis

We visualize how the agent works in our A2-RL model. We show the intermediate results of the cropping sequences, as well as the actions selected by the agent in each step. As shown in Figure 3, the agent takes the selected actions step by step to adjust the windows and chooses when to stop the process to get the best results.

We also show several cropping results of different methods on FCD [4]. From Figure 4, we can find that the A2-RL model can find better cropping windows than other methods, which demonstrates the capabilities of our model in an intuitive way. Some results also show the importance of the limited aspect ratio and history experience.





Figure 3. Examples of the sequential actions selected by the agent and corresponding intermediate results. Images are from FCD [4].

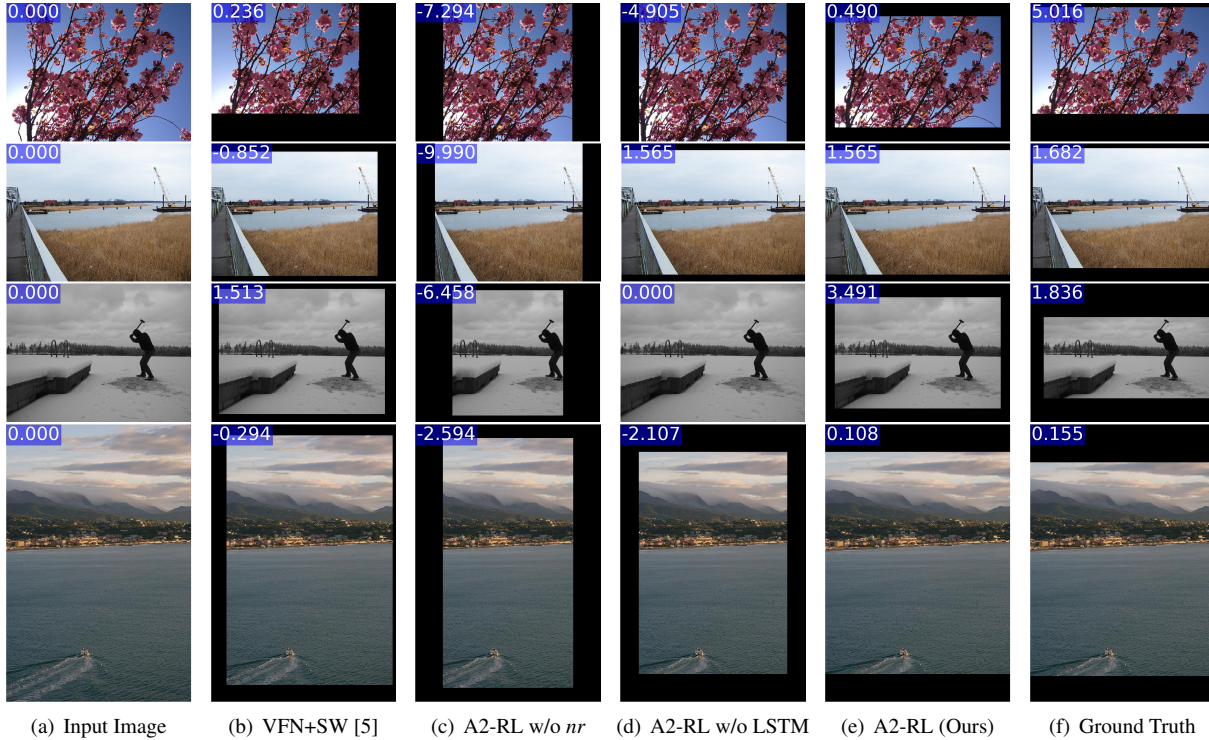


Figure 4. Image cropping examples on FCD [4]. The number in the upper left corner is the difference between the aesthetic scores of the cropped and original image, which is  $s_{aes}(I_{crop}) - s_{aes}(I_{original})$ . The aesthetic score  $s_{aes}(I)$  is used in the definition of the reward function (see Section 3.2). Best viewed in color.

## 5. Conclusion

In this paper, we formulated the aesthetic image cropping as a sequential decision-making process and firstly proposed a novel weakly supervised Aesthetics Aware Reinforcement Learning (A2-RL) model to address this problem. With the aesthetics aware reward and comprehensive state representation which includes both the current observation and historical experience, our A2-RL model learns good policies for automatic image cropping. The agent finished the cropping process within several or a dozens steps and got the cropping windows with almost arbitrary size. Experiments on several unseen cropping datasets showed

that our model can achieve the state-of-the-art cropping accuracy with much fewer candidate windows and much less time.

## Acknowledgement

This work is funded by the National Key Research and Development Program of China (Grant 2016YFB1001004 and Grant 2016YFB1001005), the National Natural Science Foundation of China (Grant 61673375, Grant 61721004 and Grant 61403383) and the Projects of Chinese Academy of Sciences (Grant QYZDB-SSW-JSC006 and Grant 173211KYSB20160008).



## References

- [1] J. C. Caicedo and S. Lazebnik. Active object localization with deep reinforcement learning. In *ICCV*, 2015.
- [2] J. Chen, G. Bai, S. Liang, and Z. Li. Automatic image cropping: A computational complexity study. In *CVPR*, 2016.
- [3] Y. Chen, Y.-J. Liu, and Y.-K. Lai. Learning to rank retargeted images. In *CVPR*, 2017.
- [4] Y.-L. Chen, T.-W. Huang, K.-H. Chang, Y.-C. Tsai, H.-T. Chen, and B.-Y. Chen. Quantitative analysis of automatic image cropping algorithms: A dataset and comparative study. In *WACV*, 2017.
- [5] Y.-L. Chen, J. Klopp, M. Sun, S.-Y. Chien, and K.-L. Ma. Learning to compose with professional photographs on the web. In *ACM Multimedia*, 2017.
- [6] 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 *ICCV*, 2017.
- [7] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV*, 2006.
- [8] Y. Deng, C. C. Loy, and X. Tang. Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine*, 2017.
- [9] S. Dhar, V. Ordonez, and T. L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *CVPR*, 2011.
- [10] S. A. Esmaili, B. Singh, and L. S. Davis. Fast-at: Fast automatic thumbnail generation using deep neural networks. In *CVPR*, 2017.
- [11] C. Fang, Z. Lin, R. Mech, and X. Shen. Automatic image cropping using visual composition, boundary simplicity and content preservation models. In *ACM Multimedia*, 2014.
- [12] E. Hong, J. Jeon, and S. Lee. Cnn based repeated cropping for photo composition enhancement. In *CVPR workshop*, 2017.
- [13] Z. Jie, X. Liang, J. Feng, X. Jin, W. Lu, and S. Yan. Tree-structured reinforcement learning for sequential object localization. In *NIPS*, 2016.
- [14] Y. Kao, R. He, and K. Huang. Automatic image cropping with aesthetic map and gradient energy map. In *ICASSP*, 2017.
- [15] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *CVPR*, 2006.
- [16] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes. Photo aesthetics ranking network with attributes and content adaptation. In *ECCV*, 2016.
- [17] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.
- [18] X. Liang, L. Lee, and E. P. Xing. Deep variation-structured reinforcement learning for visual relationship and attribute detection. In *CVPR*, 2017.
- [19] W. Luo, X. Wang, and X. Tang. Content-based photo quality assessment. In *ICCV*, 2011.
- [20] L. Mai, H. Jin, and F. Liu. Composition-preserving deep photo aesthetics assessment. In *CVPR*, 2016.
- [21] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *ICML*, 2016.
- [22] N. Murray, L. Marchesotti, and F. Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *CVPR*, 2012.
- [23] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based photo cropping. In *ACM Multimedia*, 2009.
- [24] J. Park, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon. Modeling photo composition and its application to photo rearrangement. In *ICIP*, 2012.
- [25] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.
- [26] Z. Ren, X. Wang, N. Zhang, X. Lv, and L.-J. Li. Deep reinforcement learning-based image captioning with embedding reward. In *CVPR*, 2017.
- [27] F. Stentiford. Attention based auto image cropping. In *Workshop on Computational Attention and Applications, ICVS*, 2007.
- [28] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs. Automatic thumbnail cropping and its effectiveness. In *ACM UIST*, 2003.
- [29] T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 2012.
- [30] E. Vig, M. Dorr, and D. Cox. Large-scale optimization of hierarchical features for saliency prediction in natural images. In *CVPR*, 2014.
- [31] P. Wang, Z. Lin, and R. Mech. Learning an aesthetic photo cropping cascade. In *WACV*, 2015.
- [32] W. Wang and J. Shen. Deep cropping via attention box prediction and aesthetics assessment. In *ICCV*, 2017.
- [33] R. J. Williams and J. Peng. Function optimization using connectionist reinforcement learning algorithms. *Connection Science*, 1991.
- [34] J. Yan, S. Lin, S. Bing Kang, and X. Tang. Learning the change for automatic image cropping. In *CVPR*, 2013.
- [35] L. Zhang, M. Song, Y. Yang, Q. Zhao, C. Zhao, and N. Sebe. Weakly supervised photo cropping. *IEEE Transactions on Multimedia*, 2014.