Photo Optimizing Adversarial Net with Policy Gradient Method

Ching-Yao Chuang
National Tsing Hua University
s102061145@m102.nthu.edu

Yuan-Hong Liao
National Tsing Hua University
s102061137@m102.nthu.edu

Abstract

Photo retouching enables photographers to invoke dramatic visual impressions by artistically enhancing their photos through stylistic color and tone adjustments. However, it is also a time-consuming and challenging task that requires advanced skills beyond the abilities of casual photographers. Recently, theres has been lots of tasks enhance their robustness and performance via adversarial learning. We borrow the concept of adversarial learning into photo aesthetic improving task, and use the policy gradient to optimize the policy network. To sum up, in this paper, we propose a fully automatic method, which is trained unsupervisedly, and the experiments show that our method can generalize itself to other basic filters.

1. Introduction

Image aesthetic combines brightness, contrast, tint, etc. To produce an image to meet these criteria is never a piece of cake and image aesthetic is lack of a convincing automatic evaluation matrix. All these stuff make the automatic image aesthetic enhancing stumbling.

Traditionally, people use some image processing software like PhotoShop to perform image refinement. Tuning each parameter carefully is inefficient and unscalable and for casual users, it's quite difficult to use those complicated tools. We propose a novel approach combining adversarial learning and policy gradient update to enhance photo aesthetic. Our method assume that almost all transformation on image can be decomposed into several basic image processing methods. We view these image processing as our action and agent's goal is to learn how to use the image processing method we provided to fit the discriminator's criteria. On the other hand, the discriminator's goal is to recognize which images are produced from the agent or from professional photographers.

We also study the problem of learning artistic photo enhancement styles from image exemplars. Specifically, given a set of image, each representing a photo tone and color enhancement following a particular style, we wish to learn a computational model so that for a novel input photo we can apply the learned model to automatically enhance the photo following the same style.

Figure 1. For most people, they may consider that the left images are more attractive than those in right side.

2. Related Work

Traditional image enhancement rules are primarily determined empirically. There are many software tools to perform fully automatic color correction and tone adjustment, such as Adobe Photoshop, Google Auto Awesome, and Light Room. Kaufman et al.[2] introduces an automatic method that first detects semantic content, including faces, sky as well as shadowed salient regions, and then applies a sequence of empirically determined steps for saturation, contrast as well as exposure adjustment. However, the limit of this approach is that output style is hard-coded in the algorithm and cannot be easily tuned to achieve a desired style. Yan et al.[8] cast exemplar-based photo adjustment as a regression problem, using a DNN to solve this regression problem. However their method require pair data to perform training which increases the difficulty in realistic application.

On the other hand, adversarial learning method have recently drawn significant attention. Adversarial Learning was first proposed by Lowd and Meek[5]. Goodfellow et al.[1] proposed an alternative training methodology to gen-

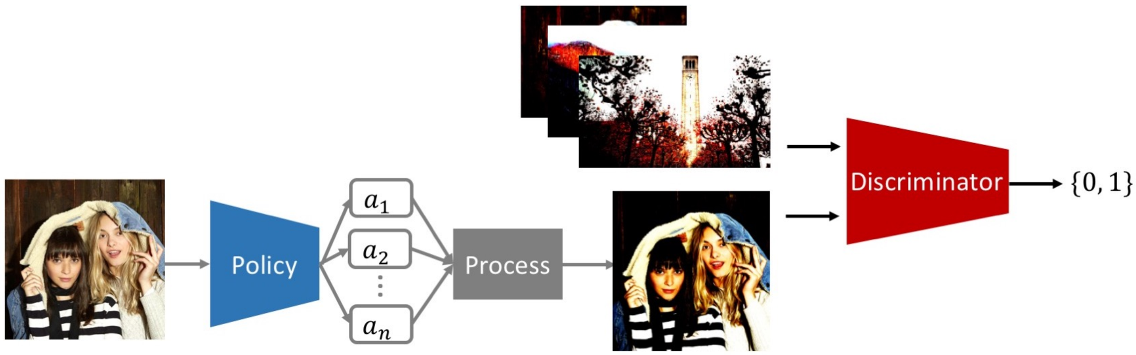

Figure 2. The architecture of the proposed model. The policy network take a photo as input and generate a set of actions (each action is for an individual feature). Then the photo will be processed based on the these actions. Discriminator will try to distinguish the processed photo.

erative models, i.e. GANs, where the training procedure is a minimax game between a generative model and a discriminative model. This framework bypasses the difficulty of maximum likelihood learning and has gained striking successes in natural image generation. In our work, we indicate that adversarial can be combined with policy gradient method to perform style transfer on photo.

3. Method

In this section, we explain our approach in more detail. First we discuss the policy gradient algorithm, which can be used to optimize any kind of reward function. Next we discuss two versions of our model with different reward giver. Instead of adversarial learning, the first version of our model receive reward from rating model pretrained on photo rating dataset, aiming to improve the aesthetic of photo. The second version of our model acquire reward from the discriminator, which can be dynamically improved in the training process. We then discuss the training algorithm, combining reinforcement learning and adversarial learning. Finally we show that our model can not only improve quality of photo but also transfer its style to designated one.

3.1. Curves

Before discussing our policy gradient method, we give a brief introduction about curve, a powerful image editing tool, which will be combined with policy. The curves tool is perhaps the most powerful and flexible image transformation, yet it may also be one of the most intimidating. Since photographers effectively paint with light, curves is central to their practice because it affects light's two primary influences: tones and contrast. Tonal curves are also what give different film types their unique characters, so understanding how they work allows one to mimic any film without ever having to retake the photograph.

The curves tool can take input tones and selectively stretch or compress them. Unlike levels however, which only has black, white and midpoint control, a tonal curve is controlled using any number of anchor points. The result of a given curve can be visualized by following a test input tone up to the curve, then over to its resulting output tone. A diagonal line through the center will therefore leave tones unchanged. In our experiments, we modify the coordinates (x and y) of N points on the curve, which correspond to policy's actions $a = \{a_1, a_2 ... a_{2N}\}$.

3.2. Photo Optimizer using policy gradient

Before the photo being fed into policy network, we first resize it to 80x80 without cropping. Then it go through three convolutional layers two fully connected layers like the architecture proposed by Mnih et al.[6] However our network is designed to predict the distribution of actions instead of Q value. For each photo, we will pick several continuous actions $a = \{a_1, a_2 ... a_n\}$, which correspond to the points on curve mentioned in section 3.1. The model is a stochastic policy $\pi_\theta(g_t|\mathbf{x})$ where $\mathbf{x}$ is the input photo and θ are the parameters of the model. Here we model curve adjusting as a continuous control task. Each action represent μ of a Gaussian distribution. In order to reduce action space, we fix the σ to a constant. While training, we will sample actions from the distribution. After acquiring actions, we will process the photo based on a, which form a generative model $G_\theta(\mathbf{y}|\mathbf{x})$ where $\mathbf{y}$ is the output photo been processed with curve mentioned in section 3.1. We directly maximize the reward $R(a)$ with the fact that

$$\nabla_\theta E_a[R(a)] = E_a[r(a)\nabla_\theta \log p(a)]$$

The details of $R(a)$ will be explained in session 3.3 and 3.4.

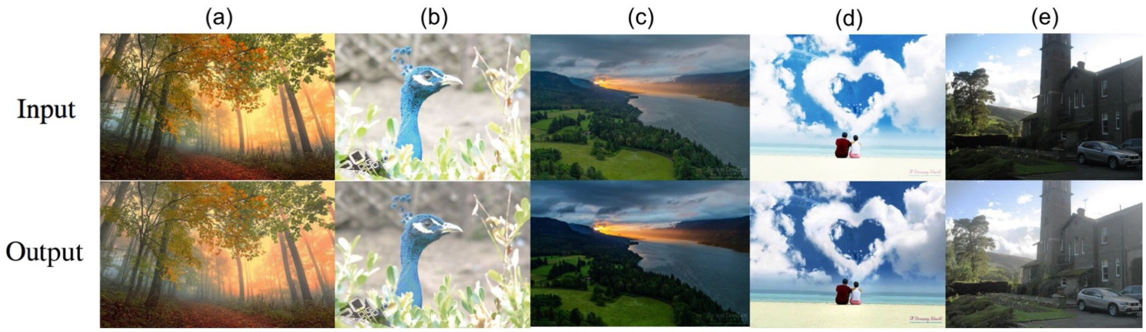

Figure 3. The top row is the model input, and the bottom row is the model output. From (b,d), the model learn to increase the lightness of the images. From (b), the model enhances the contrast of the scenery photo. From (e), the model successfully enhance the clarity of the image. Most important of all, for (a), which image is good enough, only dlittle change will be applied.

3.3. Rating Model for photo aesthetic

The first version of our model directly use the score from rating model as reward. We directly apply AlexNet[4] as our rating model and train it with Aesthetics and Attributes Database (AADB) dataset [3] which consist of 10,000 images with aesthetic score. The output of rating model is a scalar indicating the aesthetic score of input image. We set this score as the reward $R(a)$, and use the equation in section 3.2 to update the policy.

3.4. Discriminator for adversarial learning

The second version of our model acquire reward from a ϕ-parameterized discriminative model D_ϕ (Goodfellow and others 2014). D_ϕ is a probability indicating how likely a photo is well optimized, in the other word, a masterpiece like photo. The discriminative model D_ϕ is trained by providing positive examples $\mathbf{S}$ from high rating photo and the one generated by policy network. We directly use the discriminator proposed by Radford et al.[7].

3.5. Adversarial training with policy gradient

The generator model (policy) $G_\theta(\mathbf{y}|\mathbf{x})$ is to generate a photo $\mathbf{y}$ from the unprocessed one $\mathbf{x}$ so as to maximize its expected reward:

$$J_\theta = \mathbb{E}[R|\mathbf{x}, \theta] = G_\theta(\mathbf{y}|\mathbf{x}) \cdot Q^{G_\theta}_{D_\phi}(\mathbf{x}, \mathbf{y})$$

where R is the reward for a photo. Note that the reward is from the discriminator D_ϕ. $Q^{G_\theta}_{D_\phi}(\mathbf{x}, \mathbf{y})$ is the action-value function of a photo, i.e. the expected reward starting from $\mathbf{x}$, taking action a, and then following policy G_θ.

The next question is how to estimate the action-value function. In this paper, we consider the estimated probability of being a masterpiece by the discriminator $D_\phi(\mathbf{y})$ as the reward. Formally, we have:

$$Q^{G_\theta}_{D_\phi}(\mathbf{x}, \mathbf{y}) = D_\phi(\mathbf{y})$$

A benefit of using the discriminator D as a reward function is that it can be dynamically updated to further improve the generative model iteratively. It also serve as perceptual reward which means that we can grab training data from photos in the wild. Compare to Yan et al.[8], our model achieve comparable result with unpaired training data.

Algorithm 1 Photo Optimizing Adversarial Net

Require:
policy π_θ; discriminator D_ϕ; raw image dataset X; disignated image dataset Y;

```
Initialize π_θ and D_φ with random weight θ and φ;
Pre-train D_φ via minimizing the cross entropy;
repeat
    for g steps do
        Generate processed images y from x with policy
        Compute reward R(y)
        Update policy parameters via policy gradient
    end for
    for d steps do
        Combine negative examples y and positive examples S
        Train discriminator D_φ
    end for
until policy converge
```

3.6. Style Learning

Here we introduce another task called style learning. Now days there are various filters in apps like Instagram and SnapChat. Given a set of image following a particular style, we wish to learn the curve so that for a novel input photo we can apply the learned model to automatically enhance the photo following the same style.

Instead of a set of masterpiece, we replace the positive example set $\mathbf{S}$ with designated image set. For instance, we have a set of LOMO style photo, we can serve it as positive examples and apply adversarial learning. Our policy will learn to generate curve which can make the processed image deceive the discriminator. In order to fake discriminator out, the output image from policy will converge to the same style like the positive examples. Corresponding experiment result will be showed in the next section.

4. Experiments

We randomly sample 20000 images from MSCOCO train set to serve as our raw images. In photo optimizing task, we use AADB as the guide of policy network. For style learning, we random sample another 20000 from MSCOCO train set and modify its contrast with PIL library to serve as positive examples. In our experiment we fix σ to 0.5 in our experiment.

4.1. Photo Optimizing

In this section, we prove out model is content aware in. We fixed the discriminator and viewed it as an oracle criteria, which guides our policy network. In Figure 3, we can see that our model can generalize to different kind of image, including image with low/high brightness, contrast, etc. And for the images are gorgeous enough, our model will do little change on the input images. From these examples, it's obvious that our model is aware of the content in the images.

4.2. Style Learning

In style learning we dynamically train the discriminator with policy network. Here we conduct a prove of concept experiment - Learning high contrast filter. The results are showed in Figure 4. We can see that the policy successfully imitate the style of positive examples.

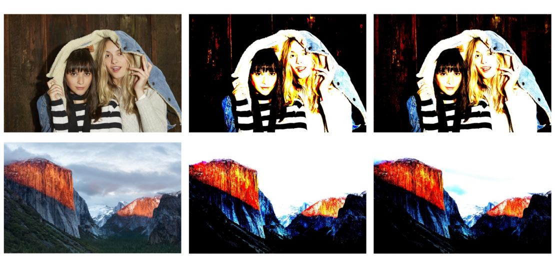

Figure 4. **Left**: Input image; **Middle**: enhanced image by our approach; **Right**: groundtruth enhanced image

5. Conclusion

We propose a brand new method for image processing, combining adversarial learning and policy gradient updates to solve non-differential problem. And our result show that our model is content aware and can potentially learn the specific style without paired training data supervision.

References

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. 2014. Advances in Neural Information Processing Systems.

[2] L. Kaufman, D. Lischinski, and M. Werman. Content-aware automatic photo enhancement, 2012. Comp. Graph. Forum 31, 8 (Dec.), 25282540.

[3] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes. Photo aesthetics ranking network with attributes and content adaptation. 2016. European Conference on Computer Vision (ECCV).

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. 2012. Advances in Neural Information Processing Systems.

[5] D. Lowd and C. Meek. Adversarial learning, 2015. International Conference on Knowledge Discovery and Data Mining (KDD).

[6] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. S. nad Ioannis Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. 2014. Nature 518, 529533.

[7] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. 2016. International Conference on Learning Representations (ICLR).

[8] Z. Yan, H. Zhang, B. Wang, S. Paris, and Y. Yu. Automatic photo adjustment using deep neural networks, 2015. ACM Transactions on Graphics.