Image Segmentation enhanced Style Transfer

Wang Ruicheng* Wang Yu* Jing Kaiyu* Liu Wei* Lin Keng Min* Wang Muqun*

* University of Science and Technology of China

Problems

In current sytle transfer, what they do is input the whole picture, then blend a style into it so that it will retains the core elements of the original image with style changed. Therefore,

- They may usually inadvertently transfer the style of the other elements(eg. background) which they don't mean to.
- The elements within the picture may interfere with each other although they don't belong to the same style, which means it's not easy to transfer different elements in the picture into different sytles.

Contributions

- A novel frame of sytle transfer is presented: image segmentation enhanced style transfer, i.e., seperate the image into different parts and then transfer them seperately so that the elements won't interfere with each other.
- In order to make the segmentation more robust, we add L0-smooth before the segmentation.
- With the segmentation results, we can try different combinations of the transfer models, each model for each element.
- Using segmentation, we can tackle with the out-of-distribution problems, which means we can better process the images that have never showed up in the training set.
- We have applied the Poisson Editing as the postprocessing to obtain the seamless combination of different transferred parts, which have boost the fidelity of our outputs.

Preliminaries

- **Image Segmentation**:
  - Our segmentation model is based on FastFCN. It uses widely used encoder-decoder structure, but adds a joint pyramid sampling(JPU) after encoder to upsample the low-resolution final feature map, getting a high-resolution feature map.
  - In detail, we use ResNet101 as encoder and Encoding head as decoder, and the whole model is trained on Pascal Context dataset.
- **CycleGAN**: Translate an image from a source domain X to a target domain Y in the absence of paired examples. Mapping $G: X \to Y$. The distribution of image from $G(X)$ is indistinguishable from the distribution Y using an adversarial loss.

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y\sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D_Y(G(x)))] \quad (1)$$

Because this mapping is highly under-constrained, couple it with an inverse mapping $F: Y \to X$ and introduce a cycle consistency loss to enforce $F(G(X)) \approx X$ (and vice versa)

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x\sim p_{data}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y\sim p_{data}(y)}[\|G(F(y)) - y\|_1] \quad (2)$$

So the full objective become:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) + \lambda\mathcal{L}_{cyc}(G, F) \quad (3)$$

where λ controls the relative importance of the two objectives, we aim to solve:

$$G^*, F^* = \arg\min_{G,F}\max_{D_x,D_y} \mathcal{L}(G, F, D_X, D_Y) \quad (4)$$

After training, what we need is G and F, which is also the style tranfer model.

- **Poisson Image Editing**
  To obtain seamless combination of outputs of several transferring models, we adopt the Poisson Image Editing at the last pass of our model. Take I_A as the result of the transferring object and I_B as the result of the transferring background, the blend image I is the solution to the partial differential equations

$$\nabla^2 I = \nabla^2 I_A, \quad I|_{\partial\Omega} = I_B|_{\partial\Omega}$$

Segmentation-enhanced Style Transfer

- Given the input X, first use L0-smooth to sharpening the major edges while eliminating a manageable degree of low-amplitude structure, then use Image Segmentation to get the mask M.
- Split X into X_1 and X_2 with the mask M.
- Process the image X or X_1 with the animal-model, which result in the output Y_1, then generate the transfered background output Y_2 with the background-model with the input X or X_2. Besides, we use the input X as the other background Y_3.
- With mask M and Poisson Image Editing, we can merge Y_1 and Y_2 or Y_1 and Y_3, to obtain the final outputs.

Experimental Results

Figure: visual comparisons of our model against the original cycleGAN: (i) original picture; (ii)&(iii) baselines(cycleGAN); (iv)&(v) our model; (vi) our model+poisson.

Figure: (ii)&(iii) baseline; (iv) our model. (v)&(vi) with poission image editing.

Future Work

- Expand ISST to more different situations, more than just animal-background picutures.
- Retrain cycleGAN and Image Segmentation with additional loss function, to get the more specific model.
- Use guided filter to get more detailed boundary in semantic segmentation, making our result more autentic.

wy2001@mail.ustc.edu.cn, lw0118@mail.ustc.edu.cn, kengmin@mail.ustc.edu.cn, proton00@mail.ustc.edu.cn, wangrc2018cs@mail.ustc.edu.cn, amoswang2000@mail.ustc.edu.cn, https://github.com/RL-WWW/ISST