Image style transform

Chen Jinye, Deng Bowen, Deng Chenguang, Qian Yicheng, Tu Zhi, Zou Muyang

Our Goal

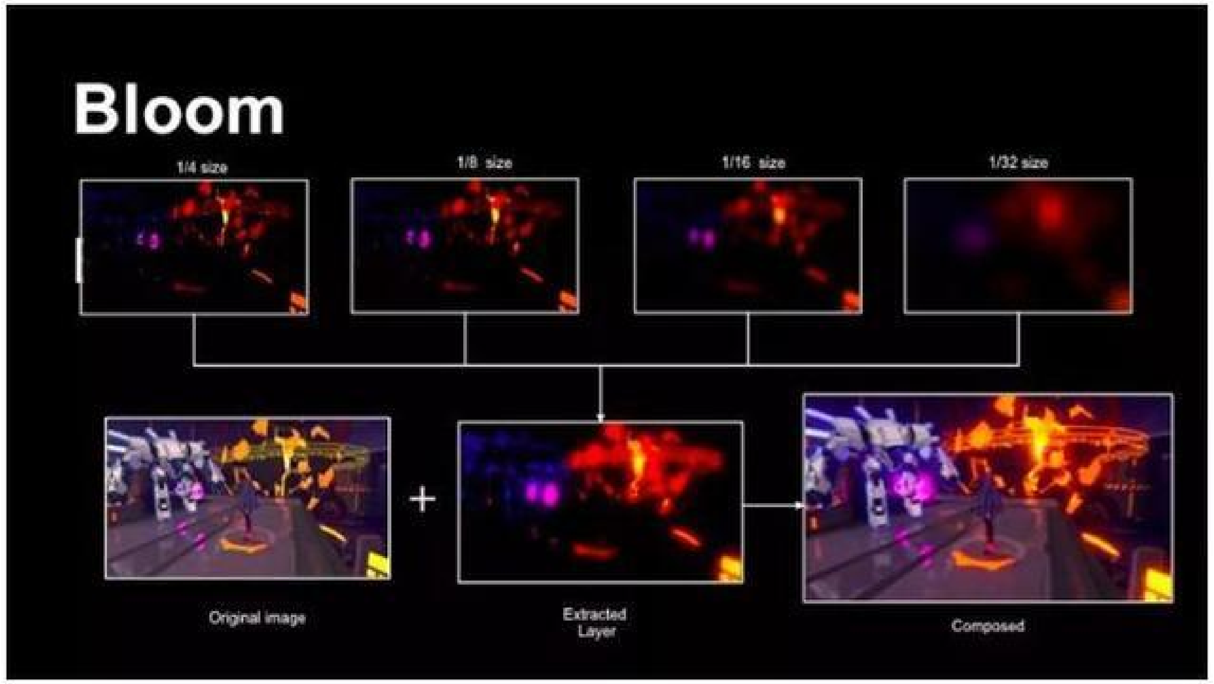
- Image Cartoonize

Related works

- Cel-shading (Cartoon rendering)
  - Filtering
  - Segmentation
- Neural networks stylization
  - depend too much on dataset
- Image synthesis with GANs
  - mostly required paired images
  - CycleGAN

Cel shading : Additional reading

- 《崩坏3》中的卡通渲染 -- 如何实现bloom效果
  - Down sampling
  - Intensity threshold
  - 4 Render target (different size)
  - Gaussian Blur
  - Superposition
  - Bloom!

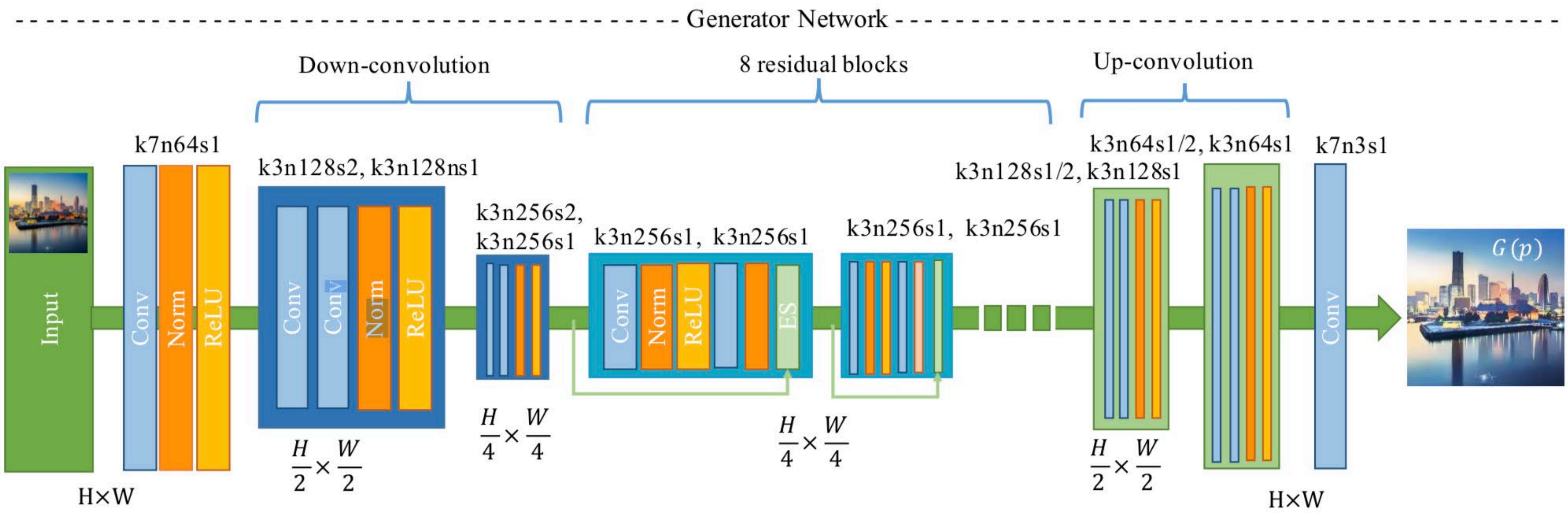


Source : http://baijiahao.baidu.com/s?id=1601902434507700874&wfr=spider&for=pc

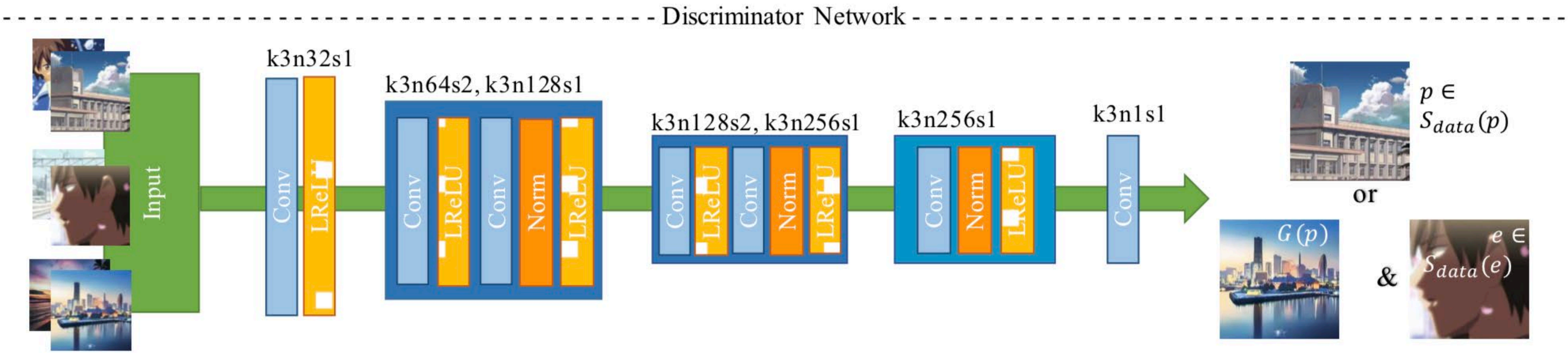
Our methods

- CartoonGAN
  - Based on: CartoonGAN: Generative Adversarial Networks for Photo Cartoonization, Yang Chen, Yu-Kun Lai, Yong-Jin Liu, CVPR 2018

Generator Network



Discriminator Network



Loss Function

$$(G^*, D^*) = \arg\min_G \max_D \mathcal{L}(G, D) \qquad (1)$$

- L(G,D) in Eq.(1) consists of two parts:
  - The adversarial loss L_{adv} (G, D)
    - which drives the generator network to achieve the de- sired manifold transformation
  - the content loss L_{con}(G,D)
    - which preserves the image content during cartoon stylization.

$$\mathcal{L}(G, D) = \mathcal{L}_{adv}(G, D) + \omega \mathcal{L}_{con}(G, D), \qquad (2)$$

Edge Promotion

(a) A cartoon image c_i (b) The edge-smoothed version e_i

- Detect edge pixels using a standard Canny edge detector
- Dilate the edge regions
- apply a Gaussian smoothing in the dilated edge regions

Adversarial loss L_{adv} (G, D)

- In CartoonGAN, the goal of training the discriminator D is to maximize:
  - The probability of assigning the correct label to $G(p_k)$,
  - the cartoon images without clear edges (i.e., $e_j \in S_{data}(e)$)
  - the real cartoon images (i.e., $c_i \in S_{data}(c)$)

$$\begin{aligned}\mathcal{L}_{adv}(G, D) = & \ \mathbb{E}_{c_i \sim S_{data}(c)}[\log D(c_i)] \\ & + \mathbb{E}_{e_j \sim S_{data}(e)}[\log(1 - D(e_j))] \\ & + \mathbb{E}_{p_k \sim S_{data}(p)}[\log(1 - D(G(p_k)))].\end{aligned} \quad (3)$$

- The second item is added to L_{adv} (G, D) of usuall GANs.

Content loss $L_{con}(G, D)$

- In addition to transformation between correct manifolds, one more important goal in cartoon stylization is to ensure the resulting cartoon images retain semantic content of the input photos.
- In CartoonGAN, we adopt the high-level feature maps in the VGG network, which has been demonstrated to have good object preservation ability.

$$\mathcal{L}_{con}(G, D) = \mathbb{E}_{p_i \sim S_{data}(p)}[||VGG_l(G(p_i)) - VGG_l(p_i)||_1] \tag{4}$$

Initialization phase

- Since the GAN model is highly nonlinear, with random initialization, the optimization can be easily trapped at suboptimal local minimum.
- To help improve its convergence, we propose a new initialization phase.
- We start the adversarial learning framework with a generator which only reconstructs the content of input images. For this purpose, in the initialization phase, we pre-train the generator network G with only the semantic content loss $L_{con}(G, D)$.

Data

- Real-world photos (7000+ images)
- Cartoon images (900+ images)

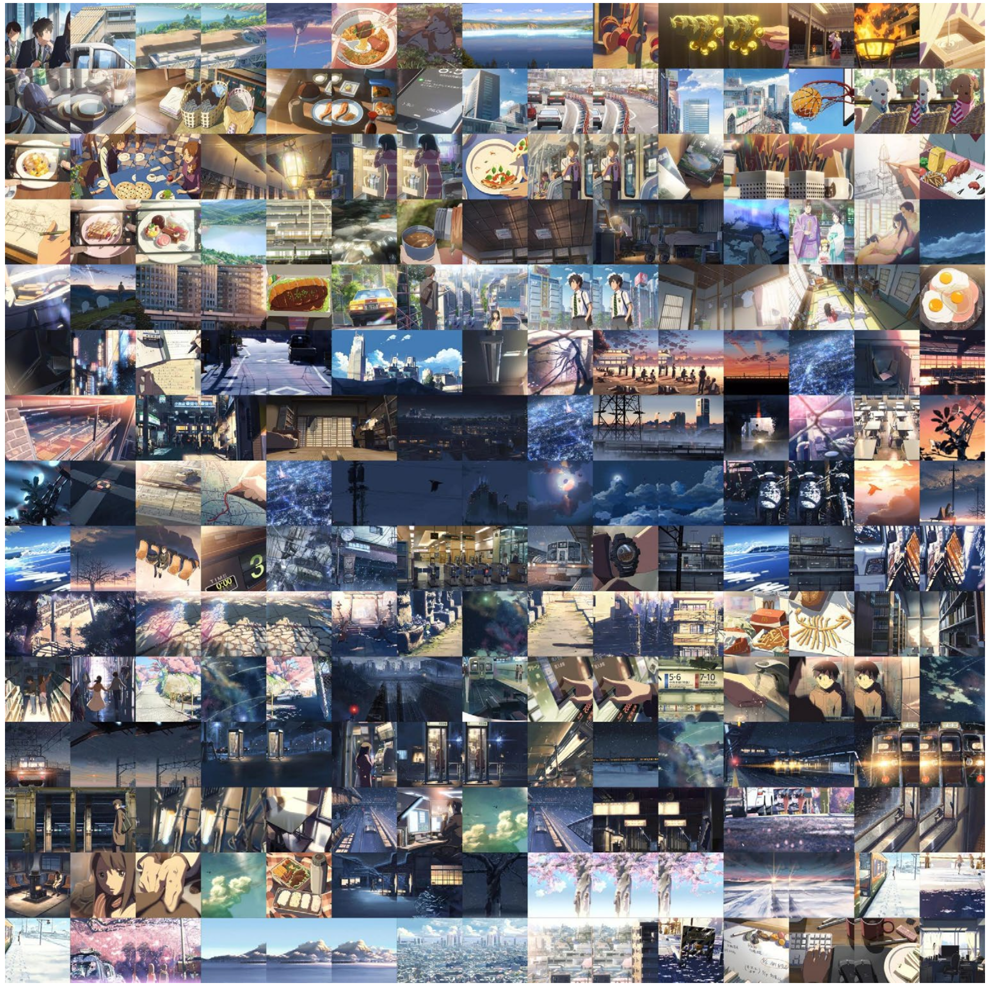
- All the training images are resized and cropped to 256×256.

Real-world dataset

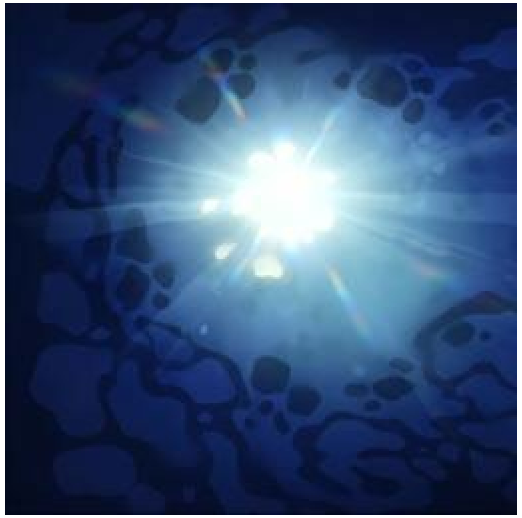
- From search engines
- From other datasets

Cartoon images

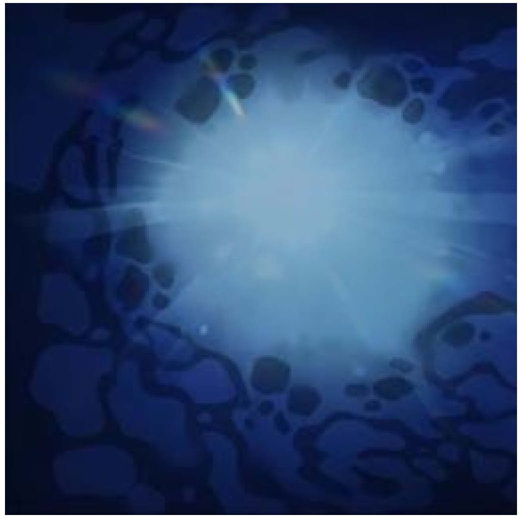
- Manually captured in cartoons



Illumination Change



Original



Illuminated

Original

Illuminated

Illumination Change

Original

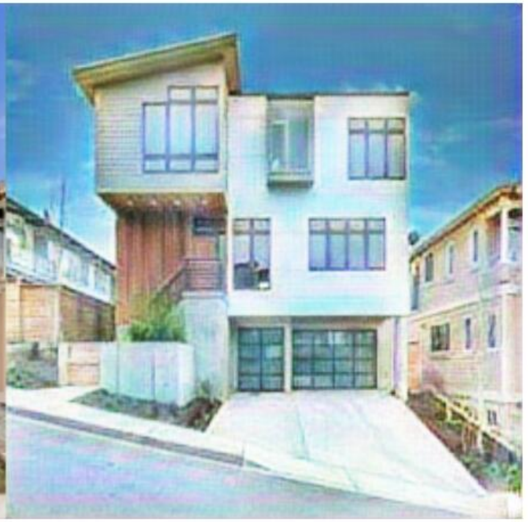
Illuminated

Original

Illuminated

Result & Discussion


Original


Cartoonized


Original

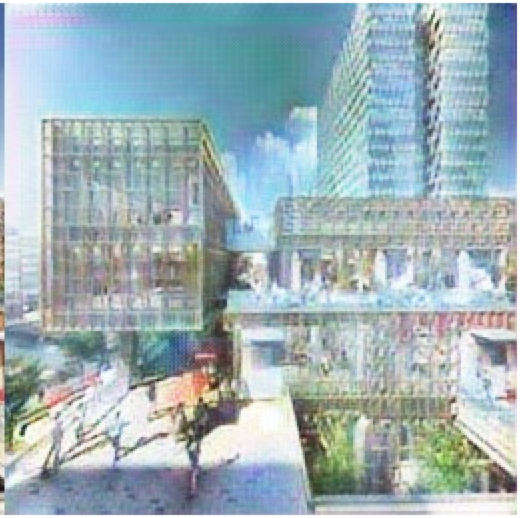

Cartoonized

Result & Discussion

Original

Cartoonized

Original



Cartoonized

Result & Discussion

Original

Cartoonized

Original

Cartoonized

Reference

- Image filtering
  - [H. Winnem oller, S. C. Olsen, and B. Gooch. Real-time video abstraction. ACM Transactions on Graphics, 25(3):12211226, 2006.]
- Formulations in optimization problems
  - [L. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via L0 gradient minimization. ACM Transactions on Graphics, 30(6):174, 2011.]
- Alternative methods rely on segmentation of images/videos
  - [J. Wang, Y. Xu, H.-Y. Shum, and M. F. Cohen. Video toon-ing. ACM Transactions on Graphics, 23(3):574–583, 2004]
- Dedicated methods have also been developed for portraits
  - [M. Yang, S. Lin, P. Luo, L. Lin, and H. Chao. Semantics-driven portrait cartoon stylization. In International Conference on Image Processing, 2010.] [P. L. Rosin and Y.-K. Lai. Non-photorealistic rendering of portraits. In Workshop on Computational Aesthetics, pages 159–170, 2015.]
- A neural style transfer (NST) method based on CNNs that transfers the style from the style image to the content image
  - [L. Gatys, A. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2414–2423, 2016.]
- Style transfer by local matching of CNN feature maps and using a Markov Random Field for fusion (CNNMRF)
  - [C. Li and M. Wand. Combining Markov random fields and convolutional neural networks for image synthesis. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2479–2486, 2016.]
- A Deep Analogy method which keeps semantically meaningful dense correspondences between the content and style images while transferring the style
  - [J. Liao, Y. Yao, L. Yuan, G. Hua, and S. B. Kang. Visual attribute transfer through deep image analogy. ACM Transactions on Graphics, 36(4):120, 2017.]
- A method to improve comic style transfer by training a dedicated CNN to classify comic/non-comic images
  - [Y. Chen, Y.-K. Lai, and Y.-J. Liu. Transforming photos to comics using convolutional neural networks. In International Conference on Image Processing, 2017.]
- pixel-to-pixel image synthesis problems.
  - [V. Dumoulin, I. Belghazi, B. Poole, A. Lamb, M. Arjovsky,O. Mastropietro, and A. Courville. Adversarially learned inference. In International Conference on Learning Representations (ICLR), 2017.]
  - [P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.]
  - [L. Karacan, Z. Akata, A. Erdem, and E. Erdem. Learning to generate images of outdoor scenes from attributes and semantic layouts. arXiv preprint arXiv:1612.00215, 2016.]
- CycleGAN
  - [J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In International Conference on Computer Vision, 2017.]
- Residual blocks
  - [K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770778, 2016.]
- Batch normalization
  - [S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, pages 448456, 2015.]

THANKS