CS172 Computer Vision I Team Project: A simple review of Image Inpainting with External-internal learning

Yufan Feng, Bin Yang, Xinyi Zhang, Dongxue Yan, Zhenxiao Yu
2019533141, 2019533230, 2019533199, 2019533144, 2019533191
fengyf, yangbin, zhangxy13, yangdx, yuzhx@shanghaitech.edu.cn

Abstract

Recent deep learning based approaches have shown promising results for the challenging task of inpainting large missing regions in an image. Most of them are directly inpainted based on color images. But sometimes color would influence the inpainting of image texture. Inspired by [8], we aim to divide image inpainting process into external and internal learning, use existing inpainting network model to inpaint gray-scale images e.g. GMCNN, Edge-Connect and DeepFill, then push this inpainted gray-scale image into colorization network to do internal learning. We will evaluate the feasibility of different models for our pipeline, at the same time change the input to study the impact of input on output.

1. Introduction

Image inpainting is a task to recover the missing regions of images and make images visually correct based on the information of the rest of images and the knowledge of image dataset. Since applications in photo processing based on image inpainting are practical and interesting, image inpainting has been widely studied in world of computer vision.

A new method called external-internel learning for image inpainting was recently proposed by professors of the Hong Kong University this year [8]. With external pretraining from large datasets and internal training on single testing images, the model uses different reconstruction nets to recover a monochromic image, i.e. a grey-scale image and then uses internel colorization technique to recover the image.

Since the new idea is innovative and effective, it is well worth giving a simple review on image inpainting through external-internal learning. In this paper, we will refer to external-internel learning and try different reconstruction networks to test their effects on grey-scale images reconstruction. In addition, we add a mask generator as an interactive tool to enrich our pipeline.

2. Related Work

A large number of approaches have emerged to solve image inpainting problems for recent years. Limited to time and detailed knowledge of neural network, we give a short review on some of the early methods.

Earlier inpainting algorithms are based on GANs, i.e., Generative Adversarial Networks [1] such as Context Encoder [7]. It comes up with the concept of 'context' and a neural network with Channel-wise fully connected layer. The every feature of the layer will be improved by every feature from the previous layer, so that the relationship of each feature will be well learned.Multi-Scale Neural Patch Synthesis [11], which is called MSNPS, is an improved method of CE. The innovation of MSNPS is a texture network. CE takes care of reconstruction and prediction, while the texture network refines the details of the filling.

Globally and Locally Consistent Image Completion [3] defines the fully convolution network with dilated convolutions which replaces the fully connected layer. In this case, GLGIC can deal with images in different scales. In addition, GLGIC also implements two discriminators of different dimensions in order that the filled images have better global and local consistency. On the basis of GLGIC, Patch-based Image Inpainting with GANs was proposed with the combination of residual learning [2] and PatchGAN [4]. The highlight of the model is to use a matrix of predicted labels instead of a single number so that it can describe the realness of the input more vividly.

Shift-Net [10] combines traditional "copy-and-paste" and modern CNN methods then proposed the shift-connection layer. The first idea encourages the decoded features of the missing parts to be closed to the encoded features of the missing parts so that a reasonable estimation of the missing part is well calculated. The second idea uses shift-connected layer to refer to the nearest neighbours of the missing parts so that we can refine both global semantic structure and local texture details.

In addition to the methods proposed above, there are some other image inpainting methods which are more advanced and more effective, such as Generative Multi-column Convolutional Neural Networks, Edge Connect, Contextual Attention and gating convolution. We will have a more detailed study in these methods and apply them into our pipeline so that we can compare the results and choose a better reconstruction model.

2.1. GMCNN

Generative Multi-column Convolutional Neural Networks [9] expands the importance of sufficient receptive fields. As shown in figure 1, the network uses a generator to produce results and globallocal discriminators for adversarial training. It includes three parallel encoder-decoder branches to extract different levels of features. The various receptive fields capture different levels of information including global and local information, which perform better on searching nearest neighbours.

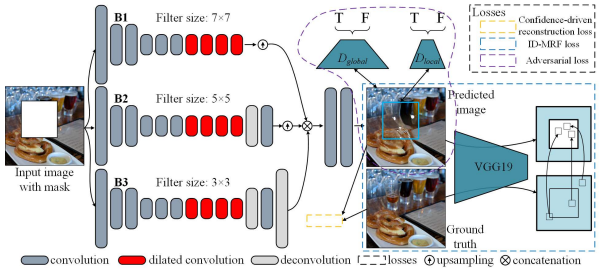


Figure 1: GMCNN network

In addition, the GMCNN network uses the Implicit Diversified Markov Random Field (ID-MRF) loss as the loss function to guide the generated feature patches to find their nearest neighbours around the missing areas as references. In this case, more local texture details can be simulated because the nearest neighbours from ID-MRF are sufficiently diverse.

2.2. Edge Connect

Edge Connect network [6] uses two encoder-decoder networks which split the inpainting process into two processes. The first network uses images with missing regions as input and the edge images as ground truth, so that the network becomes an edge generator which can recover the edge of the missing regions of the image.

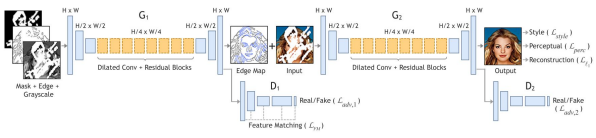


Figure 2: Edge Connect network

After the edge image is generated, the model passes the edge image and the image with missing region into the second network. With the knowledge of the local color information and the global edge information, the network learns and outputs the image with recovery of missing region. Each stage follows an adversarial model including a pair of generator and discriminator to update parameters.

2.3. Deep Fill

Free-Form image inpainting [13] inherits the network of Contextual Attention [12] which is the early version of deep filling method, so Free-Form image inpainting is also called Deepfillv2. This network proposes a new convolution method to replace the traditional convolution called gated convolution. Since the incomplete image brings some missing regions which is incorrect information, traditional convolution may take them into consideration equally. Gated convolution uses a sigmoid funtion to adjust the weight of each pixel,so the model can know the validness of each pixel.

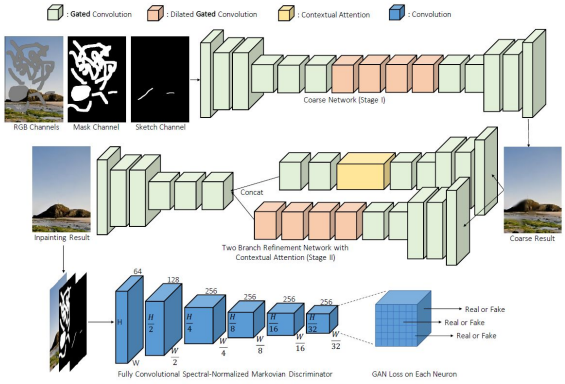


Figure 3: Overview of our framework with gated convolution and SN-PatchGAN for free-form image inpainting.

Figure 3: Deep Fill network

However, using network with gated convolution and dilated gated convolution, the output images show blurry regions. To better recovery the images, the model passes the images through the second stage of network, which contains a network using Contextual Attention (DeepFillv1) [12] and a network same as the first stage. With the same decoder, the combination of two networks in the second stage form a clear image. The model also uses SN-PatchGAN to calculate losses and enhence the training process.

3. Pipeline

Since we divide the image inpainting process into external and internal learning, we need to do some preprocess on dataset. In this section, we first implement a mask generator which allow us to draw any masks we want and generate images automatically which will be used in network later.

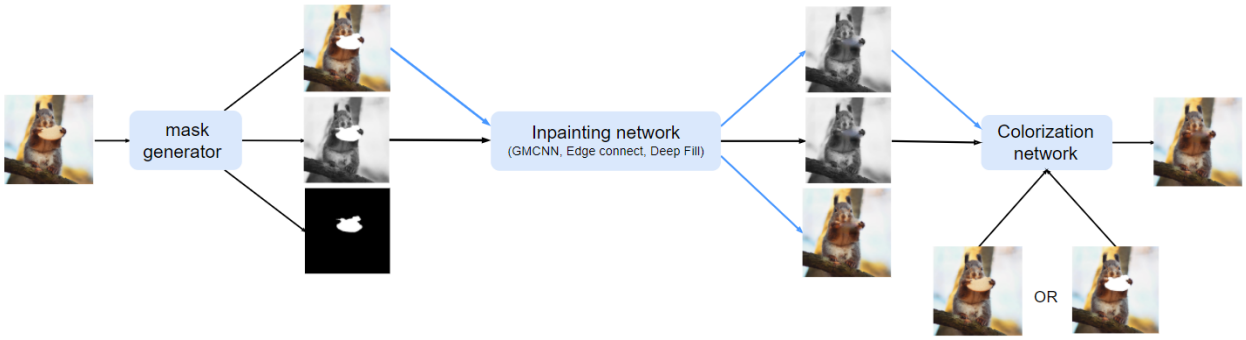


Figure 4: Pipeline

We use the three chosen inpainting network models (GMCNN, Edge-Connect and DeepFillv2) to generate gray-scale images and pass them into colorization network for internal learning.

3.1. Mask Generator

We implement a user interface that allows people to load images and draw mask themselves, then generator will generate gray-scale-mask image, color-scale-mask image and mask automatically. In our experiment, we use it to generate freeformed mask and semantic mask.

3.2. Inpainting Network

To evaluate the results of different models for our pipeline, we pass the grey-scale images and the original images with missing regions to the three reconstruction networks. To ensure performance, we use the pretrained models of GMCNN, Edge-Connect and DeepFillv2. We write a script which passes the same images to the three networks at the same time, and collect the output images for further process.

3.3. Colorization Network

Since non-missing regions usually consist of a large amount of pixels, the correspondence is extremely dense and covers most of patterns. In addition, structures in the inpainted missing region and the non-missing region are often highly correlated, some traditinal colorization fails with large regions.

The recent work [5] inspires a deep neural network to implicitly propagate color information. As long as the model learns the color mapping function in the non-missing region, we just apply it to the missing regions so that we can recover the color of the image. Considering that single grey color can map to different polychromic values, it's better to use a progressive colorization network to combine the local and global color context.

4. Experiments

4.1. Dataset

We evaluate our method on two public datasets:

Places365-Standard is the core set of Places2 Dataset, There are 1.8 million training images from 365 scene categories for training Places365 CNN. There are about 50 images for each category in the test set and about 900 images for each category in the training set.

CelebA is a large-scale face attributes dataset with over 200K celebrity images.It is widely used in face related computer vision training tasks, including face detection training, face attribute identification training and landmark marking.

Besides, we have our own photos.Based on these datasets, we put them into our mask generator, which generate gray-scale-mask image, color-scale-mask image and mask automatically. The mask can be divided into two groups: Free-Formed and Semantic masks. The former one contains a bunch of scattered dots while the latter one contains a relatively continuous block.

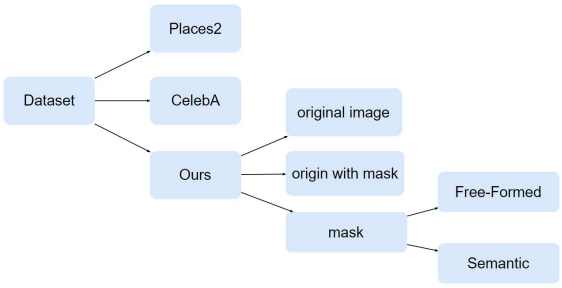


Figure 5: Dataset

4.2. Testing

In order to evaluate "Grey-Scale-image-inpainting + Colorization" model, we compare its results to nor-

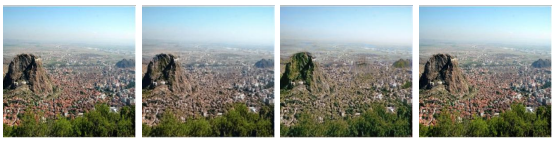
mal "RGB-image-inpainting" model and "RGB-image-inpainting + Colorization" model on Place2 dataset.

First, through the mask generator, we get color-mask-input, gray-mask-input and mask.

Second, for "RGB-image-inpainting" model, push color-mask-input into inpainting to get color-image-output and its grey semantic information. For "Grey-image-inpainting + Colorization" model, push grey-mask-input into inpainting to get grey-image-output as its grey semantic information.

Third, seperately push these two grey semantic information from two inpainting output into colorization network to reconstruct color image.

We call them briefly "G2G2RGB", "RGB2G2RGB" and "RGB2RGB".



(a) Origin (b) G2G2R (c) R2G2R (d) R2R

4.3. Evaluation

Our model is based on Places2, CelebA and our own dataset. To evaluate our output, we compare the input and output by two evaluation matrixes: peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). As the result in table1, Grey-to-RGB and RGB-to-Grey-to-RGB both perform well than RGB-to-RGB so as to the colorization. Besides, the result of Grey-to-RGB is better than that of RGB-to-Grey-to-RGB.

5. Conclusion

In this paper, we give a simple review of image inpainting with External-internal learning [8]. In our pipeline, we first implement a user interface to generate a hand-written mask and a corresponding grey-scale-mask image. By browsing some famous methods of image inpainting, we choose three advanced and effective reconstruction networks (GMCNN[9], Edge Connect[6], Deepfillv2[13]) as our inpainting network for comparison. When we pass grey-scale images or RGB images through the inpainting network, we will get corresponding grey-scale images or RGB images and pass them into our progressive colorization network. Since colorization network generates color only depending on the segmentatioon of the images, finally we will get colored images no matter we pass in RGB images or grey-scale images.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Commun. ACM*, 63(11):139–144, oct 2020.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[3] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):1–14, 2017.

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[5] Jingyuan Li, Fengxiang He, Lefei Zhang, Bo Du, and Dacheng Tao. Progressive reconstruction of visual structure for image inpainting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[6] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z Qureshi, and Mehran Ebrahimi. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212*, 2019.

[7] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[8] Tengfei Wang, Hao Ouyang, and Qifeng Chen. Image inpainting with external-internal learning and monochromic bottleneck. *CoRR*, abs/2104.09068, 2021.

[9] Yi Wang, Xin Tao, Xiaojuan Qi, Xiaoyong Shen, and Jiaya Jia. Image inpainting via generative multi-column convolutional neural networks. *arXiv preprint arXiv:1810.08771*, 2018.

[10] Zhaoyi Yan, Xiaoming Li, Mu Li, Wangmeng Zuo, and Shiguang Shan. Shift-net: Image inpainting via deep feature rearrangement, 2018.

[11] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. High-resolution image inpainting using multi-scale neural patch synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[12] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[13] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4471–4480, 2019.

Method	Grey-RGB		RGB-Grey-RGB		RGB-RGB	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
img_1	25.184	0.95239	24.858	0.94981	19.271	0.88839
img_2	23.711	0.87102	17.936	0.86937	17.445	0.80703
img_3	33.383	0.97911	34.910	0.98287	32.900	0.96802
img_4	36.043	0.97487	31.250	0.96694	30.749	0.94709
img_5	29.940	0.97249	31.014	0.97558	30.778	0.96919
img_6	29.902	0.97808	30.791	0.98210	27.370	0.95991
img_7	24.199	0.92735	23.537	0.92936	26.295	0.92228
img_8	23.791	0.94297	23.699	0.92877	20.269	0.86486
img_9	23.212	0.91200	21.561	0.90440	18.880	0.88599
img_{10}	31.397	0.97502	31.178	0.97822	30.014	0.97433
img_{11}	35.295	0.97716	34.633	0.97907	25.075	0.94480
img_{12}	25.226	0.97441	24.107	0.97197	23.780	0.96018
img_{13}	28.988	0.97742	31.368	0.97892	22.810	0.94742
img_{14}	24.535	0.94738	25.678	0.94752	23.152	0.91963
img_{15}	22.719	0.90359	21.509	0.89740	20.530	0.85523

Table 1: table of PSNR and SSIM by different method based on Place2(RGB to GRAY to RGB)