The Perfect Way to Generate “Good” Gel Electrophoresis Images

Tan Modong
moutoutan@gmail.com

※Don't try this for any research.

Abstract

Recently, there have been many deep learning based image generation method, none of them was designed for biological experiment related images. In this paper, we proposed a concept of efficient method for agarose gel electrophoresis image generation in order to skip time-wasting polymerase chain reaction (PCR) and gel electrophoresis experiment. Based on deep convolutional generative adversarial networks (DCGAN)[1], successfully generated a “good” gel electrophoresis images and destroyed the evidence ability of those images in biological paper. Our results show that a vulnerability of the evidence ability of the traditional gel electrophoresis image.

Introduction

In biochemistry, molecular biology, genetics and clinical chemistry the agarose gel electrophoresis(Fig. 1a) is used to detect and/or to separate target macromolecules(Fig. 1b) such as DNA[2].

These days, the gel images from the PCR result of polymerase chain reaction (PCR) taken by CCD camera and record as a digital data(Fig. 1). When we couldn’t detect target DNA band in the PCR image, honest person will check the protocol, DNA quality, contamination and so on to repeat experiment to get good result. In contrast to honest person, some person willing to get new “good” results without repeating time-wasting experiments by manipulation of existed PCR images.

Traditionally, copy-paste based post-processing is used to add new “good” band on the PCR image[3] but it’s very easy to detect by simple edge detection filter because copy-paste process will make a spatial discontinuity (e.g. edge-noise) on the boundary between original image and copy-pasted band image.

For example, some article was retracted[3 Fig.1i] because image manipulation and other fabrication was detected. This kind of manipulation can be detected by simple linear/non-linear image filter that enhance its spatial discontinuity.

Consequently, inappropriate post-processing that can avoid the forgery detection getting more concern in this “good” PCR image generation task.

Recently, convolutional neural network based image generation, Deep Convolutional Generative Adversarial Network(DCGAN) was developed[1]. These methods can generate various fake images. We tried to use this method to generate “good” PCR image.

Method

Architecture

I used this DCGAN code[4] to training a generation of PCR images. Only modified the resolution 28×28 to 28×56 for PCR image generation. A noise input is processed by 2 dense-(batch normalization)-Activation block. Next, reshaped dense layer output was upsampled by 2 (upsampling)-(convolution)-(activation) block to generate 28×56 PCR image.

A discriminator model is simple classifier consist from 2 (convolution)-(activation) and 1 dense-activation-dropout-dense block without batch normalization. This discriminator model classifying generator output is true PCR image or not. The true/fake PCR image was resized into 28×56 and input into discriminator model.

These generator and discriminator models were simultaneously trained to generate "good" PCR image or classifying images respectively.

Dataset

At first, downloaded 23 agarose gel electrophoresis images(PCR images) from internet. Next, concatenate images into one image for simplicity and save the concatenated images. Define crop coordinate of gel band in the concatenated images by hand using ROI manager of ImageJ and cut out each band by ImageJ macro. Finally, acquired 636 PCR images for training and validation.

Training

All network of DCGAN were trained from scratch using Adam optimizer[5] with beta_1=0.9, beta_2=0.999 Learning rate was set to 1e-5 and trained for 25045 iterations. We impremented the network in Python with Tensorflow/Keras framework[6]. All experiments were conducted under a Windows laptop with Geforce 965M, Intel Core i7-6700HQ and 8GB RAM.

Result

Randomly generated 100 images, in some case, found image with corruption or noise(Fig. 3), but there was basically no problem with generated PCR image(Fig. 4).

Conclusion

We utilized DCGAN to generate small PCR image. Sucessfully, generated small "good" quality PCR images. Our research can be used to skip time-wasting experiment step. Researcher using traditional agarose gel electrophoresis will use their time efficiently for more creative experiment not for PCR and agarose gel electrophoresis.

Many researcher can't distinguish generated PCR image is fake or real with high probability. In the future, when GAN generated gel electrophoresis image has become widespread in the field of biology, the evidence ability of the traditional gel electrophoresis image will be lost.

Futurework

We used primitive DCGAN to generate small image. Fortunately, many researcher established a GAN for high resolution image (1k~ or more). There is a way to generate full-size PCR image.

Supplementaly information

Codes are available at this repository (https://github.com/sansyo/fake_electrophoresis).

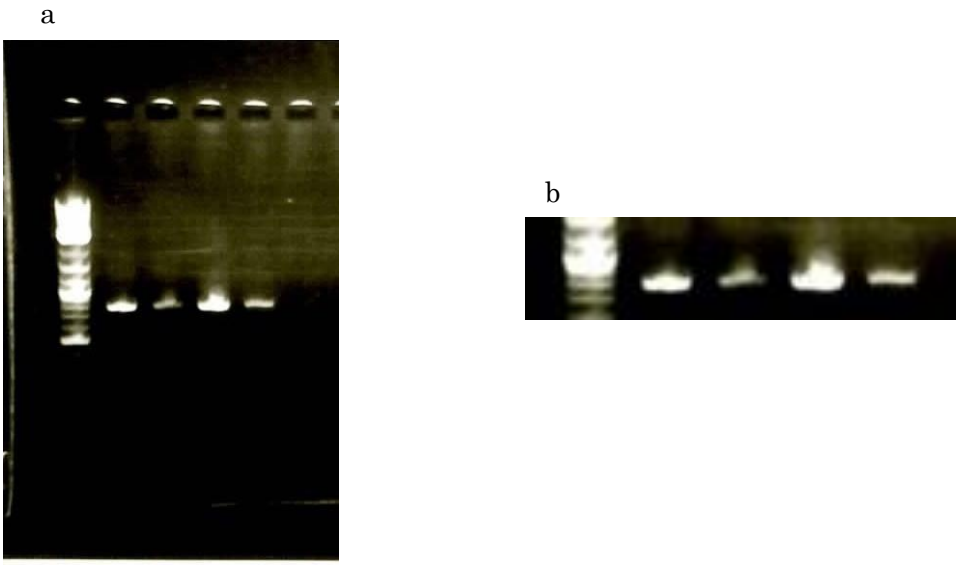

Fig. 1 REAL agarose gel electrophoresis result of PCR products

a) Example of **REAL** image of an agarose gel electrophoresis of PCR product. b) In this case, confirmed the product length is nearly 0.5 kbp by 2-Log DNA ladder.

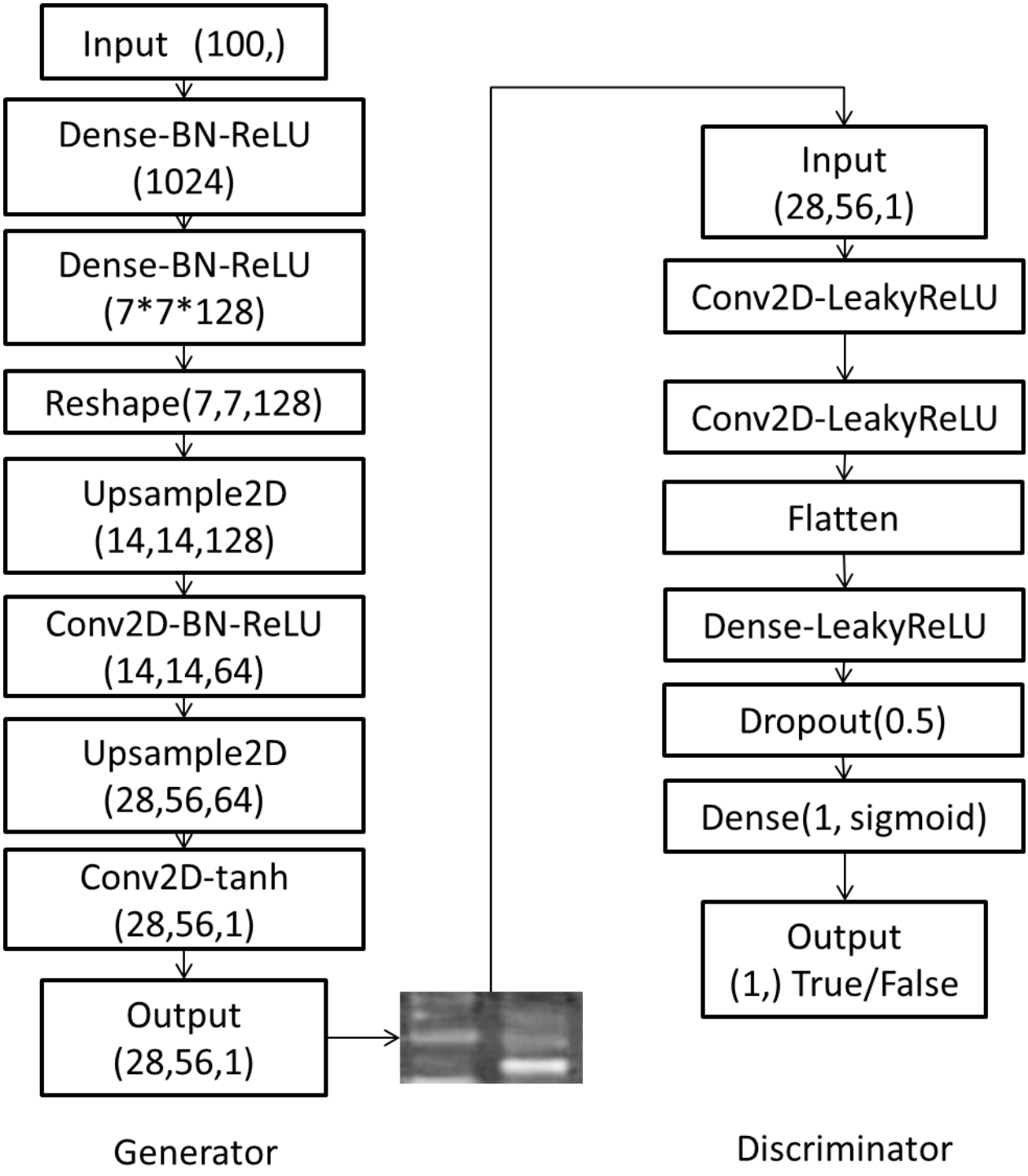

Fig. 2 Architecture of DCGAN left) Generator, right) Discriminator

Fig. 3 Corrupted PCR image sample generated by trained generator.

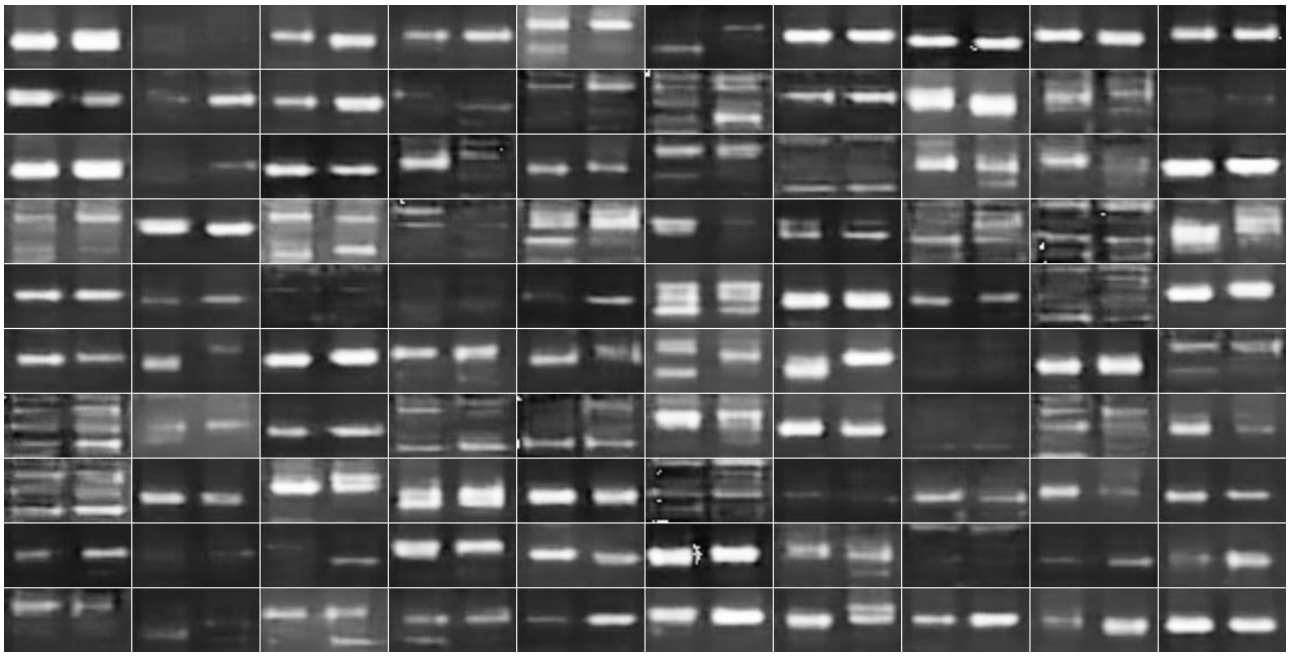

Fig. 4 Randomly generated PCR image results

References

1. Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015). https://arxiv.org/pdf/1511.06434.pdf
2. Edwards, K., C. Johnstone, and C1 Thompson. "A simple and rapid method for the preparation of plant genomic DNA for PCR analysis." Nucleic acids research 19.6 (1991): 1349.
3. Obokata, Haruko, et al. "RETRACTED ARTICLE: Stimulus-triggered fate conversion of somatic cells into pluripotency." Nature 505.7485 (2014): 641-647.
4. https://github.com/Ujitoko/keras_trial/tree/master/DCGAN
5. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
6. François Chollet and others. “Keras” https://keras.io (2015)