EVALUATION OF GAN ARCHITECTURES FOR VISUALISATION OF HPV VIRUSES FROM MICROSCOPIC IMAGES

Xiaohong W. Gao
Department of Computer Science
Middlesex University
London, UK
0000-0002-8103-6624

Xuesong Wen
Department of Nature Science
Middlesex University
London, UK
x.wen@mdx.ac.uk

Dong Li
Department of Nature Science
Middlesex University
London, UK
d.li@mdx.ac.uk

Weiping Liu
School of Electronic and Optical Engineering
Nanjing University of Science and Technology
Nanjing, China
0000-0002-9763-0732

Jichun Xiong
School of Electronic and Optical Engineering
Nanjing University of Science and Technology
Nanjing, China

Bin Xu
School of Electronic and Optical Engineering
Nanjing University of Science and Technology
Nanjing, China

Juan Liu
School of Electronic and Optical Engineering
Nanjing University of Science and Technology
Nanjing, China

Heng Zhang
School of Electronic and Optical Engineering
Nanjing University of Science and Technology
Nanjing, China

Xuefeng Liu
School of Electronic and Optical Engineering
Nanjing University of Science and Technology
Nanjing, China
liuxf1956@163.com

***Abstract*— Human papillomavirus (HPV) remains a leading cause of virus-induced cancers and has a typical size of 52 to 55nm in diameter. Hence conventional light microscopy that usually sustains a resolution at ~100nm per pixel falls short of detecting it. This study explores four state of the art generative adversarial networks (GANs) for visualising HPV. The evaluation is achieved by counting the HPV clusters that are corrected identified as well as drug treated cultured cells, i.e. no HPVs. The average sensitivity and specificity are 78.81%, 76.37%, 76.62% and 84.71% for CycleGAN, Pix2pix, ESRGAN and Pix2pixHD respectively. For ESRGAN, the training takes place by matching pairs between low and high resolution (x4) images. For the other three networks, the translation is performed from original raw images to their coloured maps that have undertaken Gaussian filtering in order to discern HPV clusters visually. Pix2pixHD appears to perform the best.**

Keywords*— *Generative adversarial network (GAN), super resolution, Human papilloma virus like particles (HPVLPs), Pix2pixHD, CycleGAN

I. INTRODUCTION

This work concerns with the determination of the presence of human papillomavirus (HPV) acquired using conventional light microscopes by employing state of the art computational imaging techniques, which constitutes being the first one to monitor HPV from light microscopic images and could potentially contribute to the development of an effective anti-HPV drugs in the future.

The human papillomavirus, or HPV, is a small, non-enveloped, and double-stranded DNA virus. A papillomavirus has a diameter of 52–55 nm [1] and infects mucosal by inducing cellular proliferation. A high risk **HPV** remains a leading cause of virus-induced cancers, mainly being discovered in cervical and head-and-neck cancers. Among those, HPV16 and HPV18 retain the two major types that account for 70% of cervical cancer cases [2, 3].

At present, the detection of HPV mainly relies on the molecular and cellular pathological evidence through labelling HPV oncogenes or oncoproteins using the approach of polymerase chain reaction (PCR), in situ hybridisation and immune-histochemical staining. This is because conventional light microscopes (e.g. Nikon C2plus Ti2 MS (Laser scan confocal, PMT)) can only depict sample structures at a maximum of ~70nm/pixel whereas an HPV sustains a size of ~50nm in diameter. In this study, a typical HPV like particles /structures (HPVLPs) retains a size of ~30-40 nm in diameter. This is because HPVLPs, viral inclusions (VIs) or virosome are viral factories where viruses assemble and replicate themselves with the support of host cellular components which could be somehow evading from the attack of host' immune responses [4]. As a result, this miniature only takes up a space of 1 to 3 pixels in an image, exhibiting difficulties to distinguish the differentiating characters when factoring into noises or artefact derived from sample preparations.

Hence, in this study, four state of the art generative adversarial networks (GAN) are investiagted towards visualization of HPV like particles. These deep learning architectures are ESRGAN [5], CycleGAN [6]. Pix2pix [7] and Pix2pixHD[8].

GAN is a deep learning neural network denoting a class of computing machines that can learn a hierarchy of features by establishing high-level features from low-level ones based on

biologically inspired human vision systems. GAN [9], designating an approach to generative modelling using deep learning methods, such as convolutional neural networks (CNN) [10]. GAN performs an unsupervised machine learning and involves automatically discovering and learning the regularities or patterns from input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset.

The ground truth is obtained using transmission electron microscope (TEM) to identify HPV like particles (HPVLP) which can achieve at a resolution at 0.5nm/px. Fig. 1 epitomises the data sets collected in this study employing three imaging microscopic modalities at varying scales, including conventional microscope and high resolution (HR) microscopy (Fig. 1(b)(c)), and TEM (Fig. 1(a)). While a HR microscope can attain a sample image at nano scale (e.g. 5nm/px), different from EM images (Fig. 1(a)), a HPVLP exhibits a challenge to apprehend from raw images as illustrated in Fig. 1(d) where those colorful pixels become unfathomable. Hence in this study, to visualize HPV, images are undertaken Fast Fourier Transform (FFT) Gaussian filtering first so that extraneous features can be cropped by filtering out less frequency signals in the Fourier space. In this paper large and small structures are constrained at 120 and 3 pixels respectively as displayed in Fig. 1(e)).

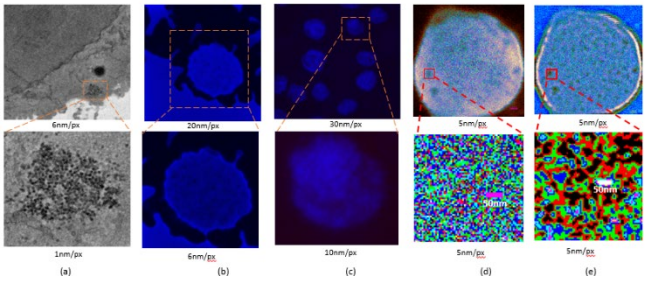

Fig. 1. Illustration of data sets applied in this study from three modalities, acquired at varying scales, including from (a): TEM, the ground truth; (b)(c): Data sets for training; (d)(e): low resolution datasets for testing. The bottom row of graphs of (a) to (c) are the selected regions on the top row of the same column that are acquired at a higher resolution.

II. Methodology

A. Dataset

In this study, as training datasets, both high resolution (HR) and low resolution (LR) images are obtained using Nikon A1plus Manual Microscope with Laser scan Confocal, GaAsP and applied. These fluorescence microscopic images (Fig. 1 (b) & (c)) were captured by scanning a microscopic slide carrying dual fluorescent labelling for HPV16 oncoprotein E6/E7 in green and nuclei in blue (CaSki Control) and HPV treated with drugs (C33a). In addition, low resolution images, applied for testing, are captured using Nikon C2plus Ti2 Microscope using conventional confocal images (Fig. 1(d)&(e)). Two band pass optical filter sets were used to acquire all these labelled cells, i.e. CaSki control. Table 1 provides details of the data acquisition information.

TABLE 1. DETAILED INFORMATION ON THE MICROSCOPIES EMPLOYED IN THIS STUDY WHERE EXW = EXCITATION WAVELENGTH AND EMW = EMISSION WAVELENGTH.

Modality	Scanning range (per pixel (px)	ExW (nm)	EmW – Red (nm)	EmW – Green (nm)	EmW –Blue (nm)	Pinhole (μm)
JEOL JEM-1400 Transmissive Electron Microscope	0.5-6nm					
Nikon A1plus Manual Microscope (Laser scan confocal, GaAsP)	5-30 nm	405, 640nm	700 (Cy5)		450 (Alexa Flour)	12.77
Nikon C2plus Ti2 MS (Laser scan confocal, PMT), (Plan Apo λ 100x Oil)	70-180 nm	488, 405nm		510-590 (FITC)	450 (DAPI)	20

While a colour image has three colour channels, i.e. RGB, all the images in this study possess colours from only two channels, either blue-red or green-red, as listed in Table 1, which are used to label HPV16 E6/E7. Hence, to form a colour image, all the images studied in this work are uniformed by converting into red and blue channels with green channels being zero, which applies to both training and testing.

B. GAN Architectures for Super Resolution Images

ESRGAN [5] is an enhanced super resolution (SR) GAN trained to transform low-resolution images (e.g. 20nm/px) to a high-resolution one (5nm/px) with four-fold (×4) increase of resolution. Specifically, different from many other super resolution network trainings where mathematical formulas, e.g. bicubic, are employed to generate low-resolution images from the available data, this work employs matching pairs of low- and high-resolution images that are scanned experimentally. Fig. 2 illustrates an the architecture of ESGAN comprising two sub-networks, a generator and a discriminator where the generator contains twenty-three Residual-in-Residual Dense Blocks.

Hence, the total loss ($\mathcal{L}$) for the generator in Eq. (1) integrate perception loss ($\mathcal{L}_{percept}$), structural similarity (SSIM) [11] loss, content loss ($\mathcal{L}_{content}$) and adversarial loss ($\mathcal{L}_G^{Ra}$).

$$\mathcal{L} = \alpha\mathcal{L}_{percept} + \gamma\mathcal{L}_{SSIM} + \lambda\mathcal{L}_G^{Ra} + \beta\,\mathcal{L}_{content} \quad (1)$$

In Eq. (1), SSIM index is defined in Eq. (2) to measure the similarity between two images in relation to spatial structure whereas the rest remain the same as applied in ESRGAN [5].

$$\mathcal{L}_{SSIM} = 1 - SSIM(x_f, x_r) \quad (2)$$

where

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{x,y} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3)$$

Where μ_x, μ_y are the averages of x, y, with σ_x^2, σ_y^2 being the variances of x, y respectively and $\sigma_{x,y}$ the covariance of x

and y. The variables of c_1 and c_2 are applied to stabilize the division when a small denominator occurs and are set to be $(0.01L)^2$ and $(0.03L)^2$ respectively, whereby L stands for the dynamic intensity range of an image, e.g. L=255 for a 8-bit image. x_f, x_r refer to fake (SR) and real (HR) images.

The α, γ, λ and β in Eq. (1) are the coefficients to balance different loss terms and are set to be 1, 0.1, 0.005 and 0.1 respectively for perceptual, SSIM, adversarial and content losses in this study.

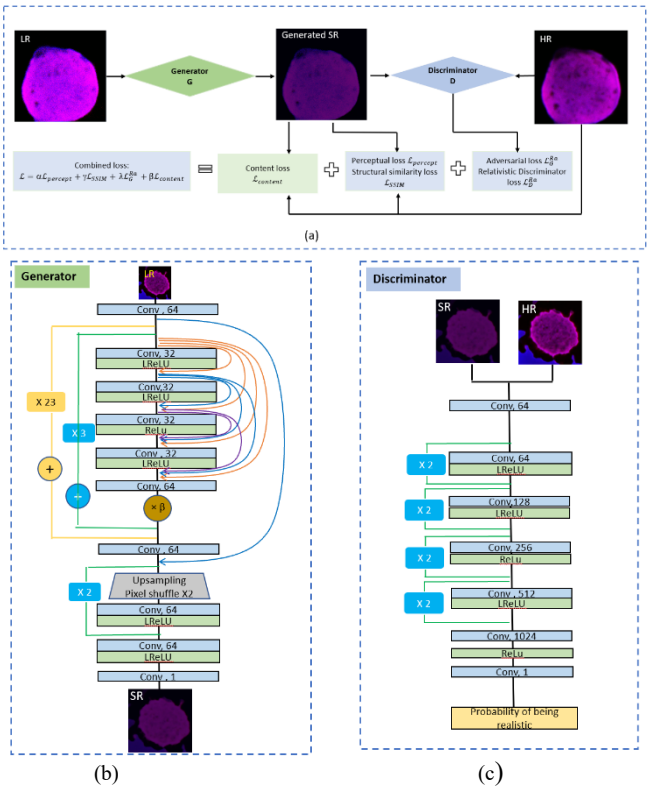

Fig.. 2. The architecture of GAN applied in this study. (a) Overall diagram. (b) Generator. (c) Discriminator.

In Fig. 3, the network of Pix2pixHD [8] is depicted, where the training takes place to translate original images (A) to its corresponding processed maps (B) undertaking FFT Gaussian filtering and being coloured using a colour lookup table.

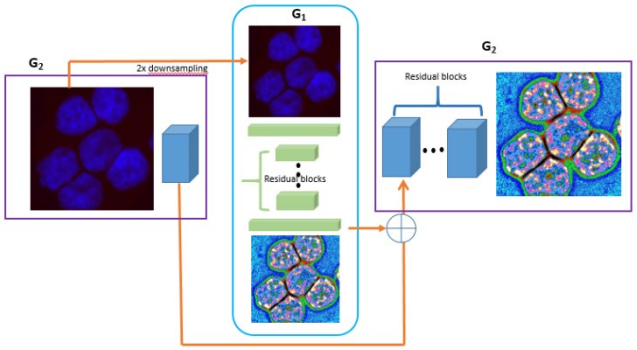

Fig. 3. The architecture of Pix2pixHD.

Fig. 3 decomposes the generator into two sub-networks, G_1 and G_2 as global generator and local enhancer networks respectively. The network first trains a residual network G_1 on lower resolution images. Then, another residual network G_2 is appended to G_1 and the two networks are trained jointly on high resolution images. Specifically, the input to the residual blocks in G_2 (right most graph) is the element-wise sum of the feature map from G_2 (left graph) and the last feature map from G_1. In this work, the input image has a resolution of 256×256 whereas the output coloured map has 1024×1024 pixels, a four-folder increase. The overall loss combines both GAN (Eq.(1)) loss and feature matching loss as formulated in Eq.(4) [8].

$$\min_G((\max_{D_1,D_2,D_3}\sum_{k=1,2,3}\mathcal{L}_{GAN}(G,D_k))+\lambda\sum_{K=1,2,3}\mathcal{L}_{FM}(G,D_k)) \quad (4)$$

Where λ is a factor that controls the importance of the above two parts, $\mathcal{L}_{FM}$ refers to the feature matching loss and D_k (k=1,2,3) the feature extractor, extracting features from each of the three blocks in Fig. 3. During the test, the parameters loadSize and fineSize are set to be 4 times of that of the input image. The Pix2pixHD will then upsample, fine tune and output a high resolution image four time as big as the input size.

In addition, the GAN-oriented networks of Pix2pix [7] and CycleGAN [6] are also evaluated. Although these two networks are designed for image translation rather than generating high resolution, they can be employed to convert raw images into coloured filtered maps to display visible HPVs. In addition, these systems do no need pixel to pixel matching pairs, the challenge facing any SR systems.

.

III. Results

The implementation is built upon Pytorch deep learning libraries [12,13], through the application of Python language. The training and testing took place under Windows 10 system with one GPU Nvidia GeForce GTX1060 with 16 Gbyte memory. The training samples comprise 2431 images with 785 for validation and are of high resolution (x4). Test samples has 100 low resolution images containing 121 HPVLP clusters with 40 samples being normal. The input size is 256×256 pixels whereas the output size from CycleGAN and Pix2pix remains the same. For Pix2pixHD and ESRGAN, the output size is 1024×1024 pixels, i.e., four time (x4) bigger. The training takes 50 epochs to complete for each network.

Table 2 lists the classification accuracy of detected HPV clusters in comparison with high resolution (4X) images of ground truth. The drug treated samples are referred as normal, which have little trace of PHVLP clusters.

TABLE 2. SENSITIVITY AND SPECIFICITY FOR FOUR APPROACHES IN PERCENTAGE (%) WHERE 'SEN'='SENSITIVITY', 'SPE'='SPECIFICITY'.

IV.	V. Sen VI. (HPV)	VII. Spe VIII. (HPV)	IX. Sen X. (Treated)	XI. Spe XII. (Treated)	XIII. AVG XIV. (%)
CycleGAN	74.28	85.98	75.40	79.60	78.81
Pix2pix	65.74	80.26	85.00	74.48	76.37
ESRGAN	70.14	81.63	77.5	77.01	76.62

ESRGAN(SSIM)	74.46	81.63	77.5	79.77	78.34
Pix2pixHD	**86.66**	**84.84**	**82.14**	**85.22**	**84.71**

To quantify the quality of generated high resolution images, spatial frequency spectrum analysis is commonly employed, which unveils the frequency extrapolation nature of the developed GAN system and the closeness of similarity between generated fake image and ground truth real image in appearance. Fig. 4 exemplifies such an example demonstrating a real (4(a)) and fake (4(b)) images and their respected frequency spectrum (in log scale) at Fig. 4(c) and 3(d). The cross-section of radially averaged power spectrum showing in Fig. 4(e) indicates an overall good agreement with largely closeness of spatial frequency spectrum.

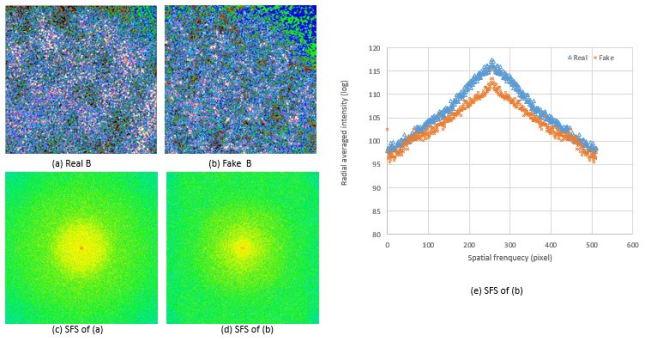

Fig. 3. Comparison of spatial frequency spectrum between generated fake image and ground truth for Pix2pixHD. (a) Real B; (b) Fake B; (c) (d) Spatial frequency spectrum (SFS) of (a) and (b) resepctively;(e) Plot of cross-section of spatially averaged power spectrum of both real (in blue) and fake images (in orange) (in log scale).

In Fig. 5, visual comparison between ESRGAN and Pix2pixHD is presented, where top row is for and bottom row for Pix2pix2D.

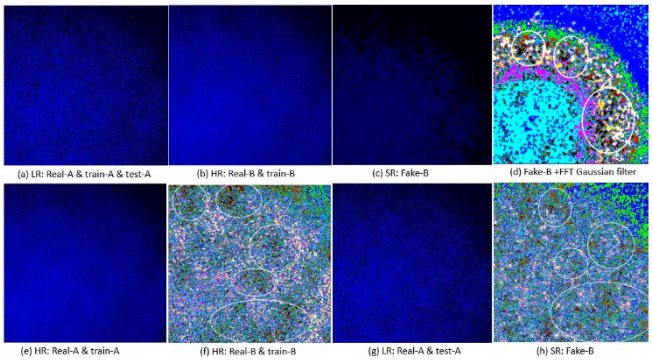

Fig. 4. Comparison the final results (last row) for ESRGAN (top row) and Pix2pixHD (bottom row) approaches. The circles are identified HPV clusters.

Because the training samples of ESRGAN are the matching LR and HR raw images, to visualize HPV clusters, the final SR image (Fig. 5(c)) is transformed by FFT (Fig. 5(d)), which appears more different to ground truth (5(f)) than 5(h) that is directly predicted by Pix2pixHD. However, since the goal of this study is to detect HPV clusters, the change of colour appearance may present no challenges. Fig. 5 also shows that by directly translating through the training from raw images to their corresponding filtered maps by Pix2pixHD can lead to more accurate results with regard to detect HPV clusters. In Fig. 5(h), 4 out of 5 HPV clusters (white circle) are detected in comparison with 5(f) (ground truth) whereas only 3 clusters in 5(d) are located.

Further qualitative comparison is presented in Fig. 6 between CycleGAN and pix2pix.

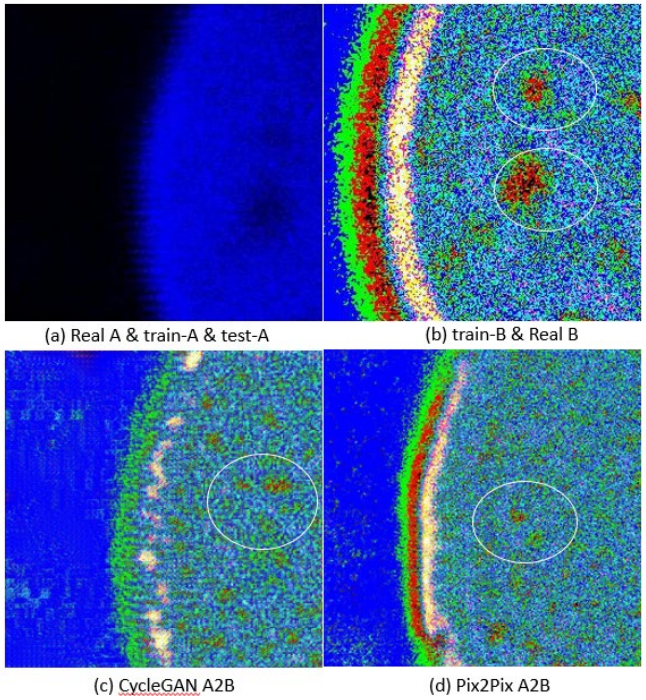

Fig. 6. Comparison results between CycleGAN (6(c)) and Pix2Pix (6(d)). White circle refers to HPV clusters.

Understandably, both CycleGAN and Pix2Pix networks are developed for image translation and only works at the same resolution between input and output. The advantage of these two networks is that they do not require paired training data. Both methods perform similar in terms of detection of HPV clusters, which can also be evidenced in Table 2.

IV. DISCUSSION AND CONCLUSION

This work might constitute the first one to identify Human papillomavirus like structures from conventional light microscopic images through the application of state-of-the-art deep learning techniques, making a step further to allow fluorescence microscopy living up to their expectations. While a conventional light microscope has established to be an essential tool in studying cell structures, super resolution appears to be the desideratum for discerning these structures at nanoscale. High risk HPVs, such as HPV16 and 18, have been confirmed to be associated with some cancers' pathogenesis, especially cervical cancers. Hence early detection of HPV can assist to identify pre-cancerous lesions that subsequently can be treated before the onset of cancer development. While detecting HPV from biological specimens is valuable in diagnosing HPV associated cancers and monitoring therapeutic effects, direct visualization of HPV or HPVLP in cells or tissue samples cannot be achieved through traditional molecular and proteomic methods. Hence microscopic imaging has been resorted to for directly observing HPV or virus particles, which calls for the increase of microscope's resolution (hence reducing pixel sizes) that

is usually above 100nm/px [14] whereas a typical HPV has a size of ~50nm in diameter. While TEM provides a solution to capture those nanostructures, it is not readily available in additional to lengthy and complex sample preparation procedures.

This study investigates four state of the art generative adversarial deep learning networks to differentiate HPV clusters under microscopy images, which are CycleGAN, Pix2pix, ESRGAN and Pix2pixHD. It has found that Pix2pixHD performs the best with sensitivity and specificity being 86% and 84% for detecting HPV clusters and 82% and 85% for detecting normal (i.e. drug treated) cells. For the other three networks, the averaged sensitivity and specificity are 78% 76% and 76% for CycleGAN, Pix2pix, ESRGAN respectively. As a result, Pix2pixHD appears to be a suitable computational imaging tool to visualize the presence of HPV clusters from LR microscopic images.

In addition to larger collection of samples for both training and evaluation, future work will ascertain the effectiveness of anti-HPV drugs in concern for treating cervical cancers or other HPV associated cancers through the employment of Pix2pixHD network to produce SR images from tissue or cellular samples. The generated SR to inspect HPV or HPVLP will provide a direct detection method for monitoring the distribution or amount of HPV load or virus like particles.

While it presents advantageous to use TEM and HR microscopy at different research centres, it limits the number of acquired datasets due to sample preparation and travelling. On the other hand, this practice has led the trained system being robust by taking in the information from different scanners, especially when the test LR datasets come from different cohort of MS scanners.

In the future, not only more cell lines infected by HPV will be included, but also the evaluation of the effectiveness of anti-HPV drugs will be conducted and verified by other molecular, cellular and proteomics laboratory techniques, raising the prospect of unravelling the insights of formation from HPV to cancer, while maintaining the prosperity of conventional light fluorescence microscopy.

Furthermore, the authors will take the findings of effectiveness of generated super resolution (×4) forward and will consider to further to increase other scales, e.g. ×8, on the premise of availability of data pairs of both LR and HR images in the future.

ACKNOWLEDGMENT

This work is financially supported by The Royal Society in the UK (Ref: IEC\NSFC\181557). Their support is gratefully acknowledged.

REFERENCES

[1] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, Human Papillomaviruses. Lyon (FR): International Agency for Research on Cancer; 2007. https://www.ncbi.nlm.nih.gov/books/NBK321770/. Retrieved in August 2020.

[2] Braaten KP, Laufer MR, Human Papillomavirus (HPV), HPV-Related Disease, and the HPV Vaccine. *Rev Obstet Gynecol*. 2008;1(1):2-10.

[3] Chelimo C, Wouldes TA, Cameron LD, Elwood JM, Risk factors for and prevention of human papillomaviruses (HPV), genital warts and cervical cancer, J. Infect, 66:207–217, 2013.

[4] Netherton C, Moffat K, Brooks E, Wileman T, A guide to viral inclusions, membrane rearrangements, factories, and viroplasm produced during virus replication. Advances in Virus Research, 70:101-182, 2007.

[5] Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Loy CC, ESRGAN: Enhanced super-resolution generative adversarial networks. In: Proc. ECCV Workshops, 2018.

[6] Zhu JY, Park T, Isola P, Efros AA, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Computer Vision (ICCV), 2017 IEEE International Conference on, 2017.

[7] Isola P, Zhu JY, Zhou, T, Efros AA, Image-to-Image Translation with Conditional Adversarial Networks, ICCV 2017, 2017.

[8] Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B, High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs, In CVPR 2018.

[9] *Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y, (2014). Generative Adversarial Networks, Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680, 2014.*

[10] *LeCun Y, Bengio Y, Hinton G, Deep Learning, Nature, 521: 436-444, 2015.*

[11] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing, 13 (4) : 600-612, 2004.

[12] Pix2pixHD, https://github.com/NVIDIA/pix2pixHD. Accessed in January, 2021.

[13] ESRGAN-Pytorch, https://github.com/wonbeomjang/ESRGAN-pytorch. Retrieved in August 2020.

[14] Klauss A, Hille C. Diffraction-Unlimited Fluorescence Imaging with an EasySTED Retrofitted Confocal Microscope, Methods Mol Biol., 1663:29-44, 2017.