cGAN basic

After the first effective implementations of GAN models, Mirza et al. introduced another effective way to generate images by applying initial conditions on generated data [1]. These additional inputs could be anything related to the data such as a particular class value or name. The generative model takes this input to generate conditional predictive results. Also, the discriminator model again has conditional acceptance to update its guess. Such models are widely used in the image-to-image translation models such as photo colorization tasks [2]. Later on, a generalized version of conditional generative adversarial network (cGAN) models are introduced with pix2pix model to learn the loss during input to output mapping. This mapping-based approach makes it possible to easily build up loss function even in very different conditions [3]. The model used in this transformation is given in the Figure 2.7.

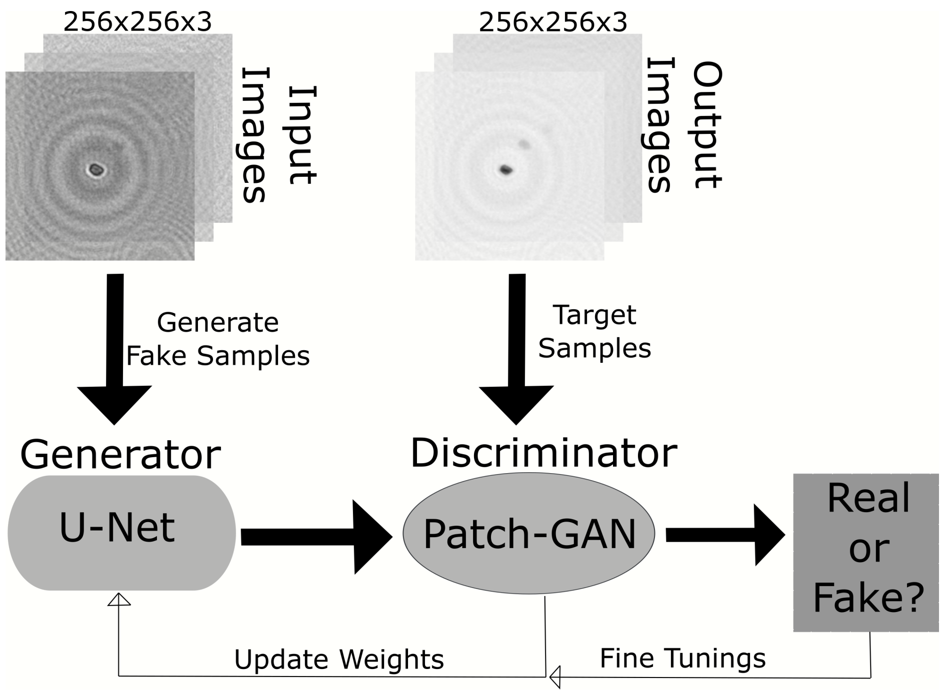

Figure 1: pix2pix conditional-GAN systematic

Reconstructed amplitude images are converted to reconstructed phase images to directly obtain arctan counterparts. In the second model, the obtained reconstructed phase images are used to convert into twin-image artifact-free counterparts that come from the single-shot iterative retrieval steps. In this way, instead of the long processing time of the one-to-one phase retrieval step, the required computation time could be eliminated by directly generating the image results with the previously trained cGAN model weights. Two models are used different datasets in the pix2pix training stage with a crop size of 256x256. During this training step, an open-source platform of ZeroCostDL4Mic [4] is used with the model of pix2pix image to image translation model [3]. In the first dataset, amplitude to phase conversion is used with directly reconstructed images. The model is trained with 1395 images. To understand the training efficiency of the model, 5 images are used to obtain mean Structural Similarity Index Matrix values (mSSIM) of images. These selected 5 images have similar intensity profiles. The mSSIM metric uses an 11 pixels window with surrounding structural similarity matrices to obtain a value between 0 to 1. It evaluates the normalized similarity between two images and a metric close to 1 represents the perfect match. During the training step, TESLA P100-PCI-E GPU is trained in 50 epochs with a patch size of 256 that is the minimal crop size. A batch size of 1 is used and the initial learning rate is set to be 0.0002 with Vanilla GAN loss function. Training takes 52 min 37 seconds.

REFERENCES

1. Mirza, M. & Osindero, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).
2. Nazeri, K., Ng, E. & Ebrahimi, M. in *Articulated Motion and Deformable Objects* 85–94 (Springer International Publishing, 2018). https://doi.org/10.1007/978-3-319-94544-6_9.
3. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. *Image-to-Image Translation with Conditional Adversarial Networks* in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, July 2017). https://doi.org/10.1109/cvpr.2017.632.
4. Von Chamier, L. *et al.* Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nature Communications* **12.** https://doi.org/10.1038/s41467-021-22518-0 (Apr. 2021).