**CV Journal Club - Pluralistic Image Completion**

GANs (Generative Adversarial Networks) [1] are widely used for various image inpainting tasks. Pluralistic Image Completion is one of the applications that is mostly used in GANs. Pluralistic Image Completion is a type of image inpainting that generates a diverse outcome to recover images by using machine learning. The issue with image inpainting is that most image completion methods only produce one result for each input. The authors’ proposed Pluralistic model will produce a wide range of possible results with different structure, color, and texture from only one image. To be precise, a single image can generate 49 different types of images. The paper’s main idea is to create two parallel paths from an input image. The first is a regenerative part, which uses the only available ground truth to obtain a prior distribution of missing parts and reconstruct the original image from the distribution. The second is a generative part in which the conditional prior is linked to the reconstructive path's distribution.

The authors split an image into three parts: Ig, Ic, and Im. Ig is an original image, Ic is a complement partial image that contains the missing pixels from the original image, and Im is a masked partial image. In other words, if we combine the clear parts (not missing) of Im and Ic, they will return to their original status. Typically, the standard image completion method is mapping Ig to Im as model training, resulting in a single image. However, the authors’ idea is taking a sample with Ic with a given Im for training model, this result will produce a variety of images. During the training, the authors use the visdom and dominate to visualize the image and loss plot. The author showed four pretrained models on their official website, one of which was trained with 24183 face images and tested with 2824 face images, both are from CelebA dataset. The authors refer to SA-GAN as their inspiration to build the generator and discriminator networks. In this paper, the ablation study is the authors use their Pluralistic model (PICNet) to compare with different methods: CVAE, “Instance Blind” and BicycleGAN. About CVAE, their output is similar to CAVE’s output since the learned conditional prior is based on the highest latent likelihood solution. Compared to “instance blind”, if visible pixels were used for reconstruction loss, the training could become poor. Also, the authors indicate there is a lack of output from BicycleGAN, causing a bad performance. As for the qualitative testing, the authors point out that quantitative analysis is difficult to test the output since one image generates several images, and the original image is no longer a benchmark against which to compare. However, they rank the top 10 samples from the discriminator, which is close to the ground truth. In this case, they have a quantitative comparison standard to do the L1 loss, peak signal-to-noise ration (PSNR), total variation (TV), and Inception Score tests.

The merit of this paper is that the authors open a new view by contributing a dual pipeline training image architecture. Because many image inpainting algorithms only can create one output if the input is one, but this paper’s algorithm can create forty-nine images and only use one image as an input, which can provide variously realistic content in the absence of missing regions [2]. Furthermore, the author overcame the qualitative testing issue, ensuring that the image output is reasonable. PICNet can be solved to large masked images, especially for large holes [3]. However, one of the flaws is that the algorithms may fool people if the damaged area is huge. In my opinion, the most difficult aspect of image inpainting is learning how to "recover" the image rather than "produce"[4] the image that is realistic-looking. In reality, sometimes the image recovering requirement is from historical images or fully destroyed images. If the PICNet overcomes the challenge of recovering images with ground truth, I believe that their model will be extensively used worldwide.

**References**

[1] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial networks. *arXiv preprint arXiv:1406.2661.*

[2] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. *In Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424. ACM Press/AddisonWesley Publishing Co., 2000.

[3] Qureshi, M., Deriche, M., Beghdadi, A., & Amin, A. (2017). A critical survey of state-of-the-art image inpainting quality assessment metrics. *Journal of Visual Communication and Image Representation, 49*, 177–191. <https://doi.org/10.1016/j.jvcir.2017.09.006>

[4] Zhang, R., Isola, P., & Efros, A. A. (2016, October). Colorful image colorization. In *European conference on computer vision* (pp. 649-666). Springer, Cham.