1. Pendahuluan

Artificial intelligence technology is currently developing very rapidly. Scientists have developed AI technology that can create images of human faces using a deep learning algorithm, known as Generative Adversarial Network (GAN). The images created by GAN look very similar to a typical human face. Even though, generated image by GAN is not a real human face. These information presented visually in Figure 1 [1].

Fig. 1. Image Illustration: (a) real face image with different resolutions, and (b) AI generated face images from left to right using various algorithms such as StyleGAN, PGGAN, Face2Face, StarGAN.

The utilization of GAN algorithm for face generation can bring positive and negative effects. One positive impact is in the business and technology sectors, such as gaming. However, there are negative impacts that can be felt by users of social media platforms, particularly in the form of identity fraud and the creation of fake accounts [2]. Many parties would be harmed if this technology were used for criminal activities. For example, someone with fraudulent intentions could create a fake account to carry out their activities. They could then find a way to falsify their photo using an image generator so that the victims would believe that the account with the AI-generated fake face is genuine. This is how artificial intelligence technology can be exploited for criminal purposes. These other gan algorithms can be found in various tools or applications. For example, FaceApp and ZAO users can use them to replace their faces with those of celebrities or other famous people, the technology is inclined to lead to DeepFake. The consequences are criminals can create fake news or false social statements that can harm many people. If the faces generated by GAN can evade face recognition systems, the security of those systems becomes highly vulnerable to threats that can cause harm [3]. Therefore, the use of AI-powered fake image detection technology is crucial in various fields, including security systems, user identity verification on social media, and forensic applications.

Prior research has been conducted by researchers in the field, and Xin Wang categorizes the detection of AI-generated fake facial images into various groups, including classification using deep learning techniques [2]. Deep learning encompasses a wide range of models and architectures. For instance, L. Minh Dang et al. employed a convolutional neural network in their research, utilizing CelebA datasets for real faces and PCGAN dataset for AI-generated faces. The approach utilized in the research involved employing a MANFA model as the architecture for a convolutional neural network. However, the utilization of the MANFA model led to a decline in performance due to imbalanced dataset manipulation scenarios. As a solution, they introduced a new model called HF-MANFA. This model reached a test set accuracy of 87.1% and have a result 93.4% of AUC value [4].

The next research is conducted by Huaxiao Mo, et al., they employed a CNN by applying a high-pass filter to the input images to create residuals, and then passing them through three groups of layers. These layers consisted of 3×3 layer convolution and 1×1 stride, accompanied by LReLu activation, resulting in 32 feature maps as output. Next, 2×2 maxpool layer and 2×2 stride was applied. Finally, the output from the last feature map were merged and fed into a fully connected layer [5]. In contrast, Songwen Mo, et al., conducted research that aimed to detect GAN-generated facial images with different color space channel combination. In order to provide a channel focus mechanism on the model's contributing characteristics at the shallow level, they were evaluating the variations in color-space components. The result of using color space channel combination, specifically H, S, and Cb, the accuracy of the test set reached an impressive 99.10% [6]. Meanwhile, the research conducted by Beijing Chen et al., focused on detecting GAN-generated faces using a modified Xception model. They proposed a method by removing four residual blocks to avoid overfitting. Additionally, they applied an inception block with dilated convolution to generate multi-scale features [7].

In this research, we developed a system to detect generated face images using deep learning with the Xception model to achieve significant accuracy.