

Article

Enhancing X-ray Security Image Synthesis: Advanced Generative Models and Innovative Data Augmentation Techniques

Bilel Yagoub ¹, Mahmoud SalahEldin Kasem ^{1,2} and Hyun-Soo Kang ^{1,*} 

¹ Department of Information and Communication Engineering, School of Electrical and Computer Engineering, Chungbuk National University, Cheongju-si 28644, Republic of Korea

² Department of Multimedia Systems, Faculty of Computers and Information, Assiut University, Assiut 71526, Egypt

* Correspondence: hskang@cbnu.ac.kr

Abstract: This study addresses the field of X-ray security screening and focuses on synthesising realistic X-ray images using advanced generative models. Insufficient training data in this area pose a major challenge, which we address through innovative data augmentation techniques. We utilise the power of generative adversarial networks (GANs) and conditional GANs (cGANs), in particular the Pix2Pix and Pix2PixHD models, to investigate the generation of X-ray images from various inputs such as masks and edges. Our experiments conducted on a Korean dataset containing dangerous objects relevant to security screening show the effectiveness of these models in improving the quality and realism of image synthesis. Quantitative evaluations based on metrics such as PSNR, SSIM, LPIPS, FID, and FSIM, with scores of 19.93, 0.71, 0.12, 29.36, and 0.54, respectively, show the superiority of our strategy, especially when integrated with hybrid inputs containing both edges and masks. Overall, our results highlight the potential of advanced generative models to overcome the challenges of data scarcity in X-ray security screening and pave the way for more efficient and accurate inspection systems.

Keywords: data augmentation; X-ray security image; generative adversarial networks; image synthesis; deep learning



Citation: Yagoub, B.; Kasem, M.S.; Kang, H.-S. Enhancing X-ray Security Image Synthesis: Advanced Generative Models and Innovative Data Augmentation Techniques. *Appl. Sci.* **2024**, *14*, 3961. <https://doi.org/10.3390/app14103961>

Academic Editor: Thomas Lindner

Received: 21 March 2024

Revised: 2 May 2024

Accepted: 4 May 2024

Published: 7 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

X-ray technology plays a crucial role in enhancing the security of public transportation systems. By penetrating luggage and other items under inspection with X-rays, these inspection systems can reveal their contents without resorting to opening them, making these systems an essential tool for finding prohibited or dangerous items hidden within. X-ray machines operate by emitting X-rays, a type of electromagnetic radiation, which pass through objects and are captured by a detector on the opposite side. Materials absorb X-rays differently based on their density and composition, creating detailed images that reveal much about an object's internal structure and contents [1]. The capacity of X-ray technology to visualize the internal composition of items without the need for direct manual examination makes it essential for security purposes.

This technology ensures safety by preventing various threats from entering secure areas, such as airports, train stations, and subway systems, thereby protecting passengers and infrastructures. X-ray technology significantly increases the accuracy of inspection processes. Unlike manual inspections, which can be time-consuming and sensitive to human error, X-ray systems provide a fast and reliable way to check the contents of luggage and the objects under inspection. They produce detailed images that allow security personnel to quickly identify suspicious items, such as weapons and explosive materials, ensuring threats are intercepted before posing a risk to public safety [2,3].

In recent applications of deep learning to X-ray imagery for security screening and medical diagnostics, researchers have utilised advanced techniques like denoising and

super-resolution to enhance image quality [4,5]. These techniques enable the extraction of more detailed information from X-ray images, leading to better diagnostic accuracy and efficiency. By employing sophisticated algorithms such as convolutional neural networks (CNNs), deep learning provides a more precise detection [6,7] and classification of features within X-ray images than traditional methods can achieve. Multitask contrastive learning can also be employed for automatic X-ray diagnosis. Innovative frameworks have been introduced to distinguish COVID-19 from other pneumonia types through X-ray analysis [8]. Strategies to enhance deep learning model detection accuracy and address complex challenges in training and testing include modifying activation functions in deep CNNs, employing transfer learning [9,10], utilising image inpainting [11,12], and applying models to tasks such as cancer diagnosis, detection [13], and classification, material discrimination [14], medical question-answering [15,16], and software engineering applications like optimizing project schedules, customer segmentation [17,18], and IoT intrusion detection [19,20]. Unique approaches, such as creatively combining activation functions and optimization systems, contribute to the advancement of deep learning models. As a result, deep learning models have become integral in various aspects of computer vision.

In the medical sector, this technology facilitates the accurate diagnosis of diseases by recognizing subtle patterns indicative of various health issues, thereby enhancing patient care with timely and precise diagnoses [21,22]. In security, deep learning automates the detection of prohibited items in scanned objects, considerably boosting both the accuracy and speed of security screenings. These advancements demonstrate the critical role of deep learning in refining X-ray imaging techniques, making it a crucial tool in improving healthcare outcomes and enhancing security measures.

In the realm of security, deep learning applications are becoming more common. Convolutional neural network (CNN)-based techniques, including Faster Region-Based CNN (Faster RCNN) [23] and You Only Look Once version (YOLOv) [24] have been adapted for detecting firearms in X-ray images. Other studies have evaluated the performance of Fast RCNN and single-shot multibox detector (SSD) [25] in both single and multiview X-ray imagery. Additionally, transfer learning approaches [1] have been explored to see if algorithms can generalize across different types of scanners in a multi-class detection problem [26–28].

Image synthesis in deep learning involves creating new synthetic images using techniques such as generative adversarial networks (GANs) or variational autoencoders [29–31]. This method is critical in fields such as medical imaging and security, where diverse and comprehensive datasets are required to train effective models. In medical X-rays, image synthesis enables researchers to expand training datasets and address the challenge posed by the limited availability of real patient data, which is often restricted due to privacy and security regulations. This expansion helps improve the models' ability to identify subtle disease indicators that are not well represented in available datasets, enhancing the performance of diagnostic tools. In security applications, especially X-ray screening, synthetic images play a critical role in enhancing the effectiveness of detection systems. The availability of X-ray datasets for security purposes is typically limited due to strict privacy and security regulations. This restricts the publication and use of real X-ray images. Researchers can simulate a wide range of security scenarios and objects, such as hidden weapons or other contraband, by generating synthetic X-ray images. This allows for the training of more robust models capable of identifying potential threats more accurately. Enhanced training through synthetic imagery not only improves the detection capabilities of security systems but also enhances their reliability, ensuring they perform effectively under diverse and challenging conditions.

Our research focuses on the application of deep learning techniques to the field of X-ray image synthesis, particularly using the Pix2Pix and Pix2PixHD models. We explore these models for their ability to generate synthetic X-ray images, a crucial capability for expanding the available data in security imaging. Unlike traditional X-ray imaging, which relies on a direct analysis of captured images, our approach involves creating new images

that are similar to real X-ray scans. The limited availability of X-ray data, often restricted due to privacy concerns in this field, presents challenges that this method addresses. Deep learning has also enhanced the accuracy of X-ray security imaging techniques. The classification of X-ray security imaging methods includes classical image analysis, image enhancement, threat image projection, and deep learning [32,33].

In this study, we use a dataset that includes a wide range of hazardous products sorted into 35 distinct categories. This dataset was created using advanced X-ray scanning technology. Through a series of carefully designed experiments, we evaluate the ability of these models to transform mask and edge data into detailed X-ray images. We compare their performance and highlight notable enhancements in image quality and realism, demonstrating the potential of deep learning to improve the accuracy of X-ray security imaging techniques.

Our study provides a comprehensive overview of our methodology, from dataset preparation and model configuration to the experimental setups and evaluation criteria used to assess performance. By presenting a comparative analysis of Pix2Pix and Pix2PixHD in the context of X-ray image synthesis, we aim to contribute valuable insights to the ongoing discourse in image-to-image translation research and its practical applications, setting the stage for future advancements in this promising field. The contributions of this study can be summarized as follows:

- We introduce a novel hybrid edge with a mask input method for enhancing the quality of generated X-ray images.
- We conduct an analysis to compare the results given by the Pix2Pix and Pix2PixHD models using various types of inputs, providing critical insights into their respective capabilities.
- We investigate how different input types affect the quality of synthesised X-ray images, highlighting the connection between input characteristics and image fidelity.
- We train the Pix2Pix and Pix2PixHD models using a comprehensive Korean dataset to generate X-ray images, highlighting their adaptability and efficiency.

This paper is organized as follows: Section 2 examines previous work in the field, concentrating on methodologies and developments relevant to our study. Section 3 details our unique Korean dataset used for the experiments, its composition, and categorization. Section 4 describes our methodology, including the use of the Pix2Pix and Pix2PixHD models, and the novel three-channel input technique. Section 5 outlines the experimental setup and presents the results, emphasizing the effectiveness of our approach through a comparative analysis. Finally, Section 6 summarizes our conclusions and proposes future directions for advancing X-ray image synthesis.

2. Related Work

In the field of X-ray security screening, the lack of sufficient image datasets for training convolutional neural networks (CNNs) present a major challenge. Addressing this issue, Yue Zhu et al. [34] made a noteworthy contribution by proposing an innovative data augmentation method for enhancing X-ray image datasets. Their research began with the generation of a diverse range of prohibited item images using an improved self-attention generative adversarial network (SAGAN). Subsequently, a CycleGAN-based technique was applied to convert natural images of items into X-ray-style images, significantly enriching the diversity in terms of item shape and pose. These images were then synthesised with background images to create realistic X-ray security checking scenes. The effectiveness of this augmented dataset was validated using two single-shot multibox detector (SSD) models, and the findings revealed that the performance of the SSD model trained with the augmented dataset surpassed that of the model trained with the original dataset. Yue Zhu's work not only tackled the critical issue of insufficient training data in X-ray security screening but also highlighted the potential of advanced generative models in generating diverse and realistic training datasets.

Also, Jinfeng Yang [35] proposed a novel data augmentation technique using generative adversarial networks (GANs). This process involved extracting prohibited items from X-ray images using a K-Nearest Neighbour (KNN) matting scheme, categorizing them based on poses estimated through a space rectangular coordinate system, and then generating realistic images with an improved GAN model. The effectiveness of these generated images was validated through a cross-validation scheme using a CNN model, showing their successful classification and indicating their potential to augment the existing dataset for CNN training. The study also focused on optimizing the GAN model, including adjustments to the architecture, loss function (CT-GAN), and specific parameter settings to enhance the quality of generated X-ray images.

Jk Dumagpi et al. [2] suggested exploring the use of generative adversarial networks (GANs) for image augmentation to improve the performance of these algorithms on an imbalanced X-ray dataset. Specifically, they employed Deep Convolutional GAN (DCGAN) for generating new X-ray images of threat objects and CycleGAN for translating camera images of threats into X-ray images. These synthesised images were then combined with background X-ray images to augment the dataset. The authors trained various Faster R-CNN models using different augmentation approaches, including both image transformation and image synthesis and assessed their performance on a practical X-ray image dataset. The results demonstrated that image synthesis effectively reduced the false-positive rate by up to 19.9% when combined with conventional image augmentation while maintaining a high true-positive rate of about 94%. This study highlighted the potential of GAN-based image synthesis in addressing the imbalance problem in X-ray security image datasets.

D Liu [36] tackled the scarcity of pseudocolour X-ray image datasets for prohibited item detection in security inspections by introducing a data augmentation method, DA-PIX. This approach utilised a generative adversarial network model, SCAB-XWGAN-GP, to generate high-quality dual-energy X-ray images. These synthetic images were then composited with real data to simulate realistic overlapping conditions, significantly enhancing dataset robustness, and improving the performance of object detection models in identifying prohibited items in X-ray scans.

Y Zhu et al. [37] proposed an innovative method to augment the dataset of X-ray images of prohibited items. They utilised an improved self-attention generative adversarial network (SAGAN) for this purpose. Initially, they collected a preliminary set of X-ray images representing various prohibited items. Key features such as colour, contour, and texture were extracted from these images to minimize background interference. The SAGAN model was then enhanced by deepening its network structure and adjusting loss functions to handle the small initial dataset effectively. The focus was on generating realistic X-ray images. An evaluation using GAN-train, GAN-test metrics, and the FID score demonstrated that the model successfully enlarged the image dataset and improved data quality for training more accurate automatic detection models.

Z Zhao et al. [38] introduced a method to enhance the detection of prohibited items in X-ray baggage scans using advanced machine learning techniques. They proposed a novel image generation approach utilising generative adversarial networks (GANs), specifically through a custom CT-GAN model, to produce realistic X-ray images of prohibited items. This method allowed for increased diversity and quantity in training datasets, essential for training effective convolutional neural networks (CNNs). By improving the CGAN model to vary the poses, positions, and scales of the generated images, they further refined the training process.

F Shao et al. [39] developed a foreground and background separation (FBS) framework for X-ray prohibited item detection, addressing the challenge of object overlap in X-ray security screening. Their method involved separating prohibited items from irrelevant objects by designing a target foreground and using recursive training to fine-tune the detection. An attention module was also integrated to enhance focus on the foreground. The framework was tested on synthetic and public datasets, showing that it significantly

outperformed existing solutions and required only a minimal number of foreground and background ground truths for effective training. This approach offered a more accurate and efficient solution for detecting prohibited items in complex X-ray images.

L Qiu et al. [40] proposed a novel method for waste inspection using X-ray imagery, addressing the inefficiency of traditional manual or RGB image-based methods. They introduced the concept of instance-level waste segmentation in X-ray images and created a comprehensive dataset of over 5000 X-ray images with detailed annotations. To tackle the challenges of heavy occlusions and penetration effects typical in X-ray images, they developed an innovative instance segmentation method, Easy-to-Hard Instance Segmentation Network (ETHSeg), which incorporated a global structure guidance module and an easy-to-hard disassembling strategy. This approach significantly enhanced the ability to segment and analyse overlapped objects in X-ray waste images, demonstrating the potential to revolutionize efficient waste inspection processes.

Table 1 outlines key generative models used in X-ray image augmentation; the table highlights differences in model types, publication years, and datasets used.

Table 1. Comparative analysis of X-ray security image generative models.

Model	[Ref.]	Year	Dataset
Custom CT-GAN	Z Zhao [38]	2018	Private dataset
GAN and KNN	Jinfeng Yang [35]	2019	GDX-ray
SAGAN and CycleGAN	Yue Zhu [34]	2020	GDX-ray and SIXray
DCGAN and CycleGAN	Jk Dumagpi [2]	2020	SIXray
Improved self-attention GAN (SAGAN)	Y Zhu [37]	2020	Private dataset
SCAB-XWGAN-GP	D Liu [36]	2022	Private dataset

3. Dataset

In the field of image synthesis and data augmentation for X-ray imaging, the dataset is the most important factor in producing high-quality images for various applications, such as detection [41] and layer separation [42]. These applications require complex datasets to successfully identify hidden prohibited images and recognize prohibited objects occluded within X-ray images. Although it is challenging to obtain such specific datasets, deep learning enables the creation of complex datasets and their associated ground truths. However, the effectiveness of these synthesised images still heavily depends on the availability of a real dataset that includes diverse objects in multiple positions and of high quality. However, there are two popular X-ray datasets: GDX-ray [43] and SIXray [44]. GDX-ray consists of greyscale images, whereas pseudo-colour images are needed for image synthesis in X-ray imaging. SIXray has over a million images, but only 8,929 of them contain prohibited items.

Due to the limitations of existing datasets, our experiment utilised a Korean dataset [45], which contained 35 types of hazardous products relevant to ports and airports. There were 570 different types of objects in this collection, all carefully categorized into 35 different groups. The dataset was created using four different scanning techniques, and the results of each technique were carefully categorized into separate files: “Single Default”, “Single Other”, “Multiple Category”, and “Multiple Others”. This structured approach ensured the creation of a comprehensive and diverse dataset, significantly enhancing its inclusiveness and practical applicability.

Single Default: This dataset captured a single sample of the relevant item, excluding any other items. The item was scanned from various angles. The data set represented 15% of the total data for the target image, considering different angles and entry directions of the luggage.

Single Other: This dataset included a relevant item captured from various angles, along with general non-hazardous objects such as clothes, cables, chargers, pens, and glasses. However, these additional items were not classified as part of our specific object

classes. The dataset represented 15% of the total data for the target image, considering the longitudinal direction.

Multiple Categories: This dataset captured a single instance of the relevant item alongside other hazardous objects. These images constituted the majority of the dataset, accounting for 40% of the total data. Simultaneous images were generated with other samples of the same item. For instance, if the target item was a gun, the images could also include other sharp tools or hazardous items at the same time.

Multiple Others: This dataset comprised multiple instances of the relevant item alongside general non-hazardous objects. These images constituted 30% of the entire dataset.

All four types of data were generated by inserting the items into bags, backpacks, suitcases, and baskets as shown in Figure 1.

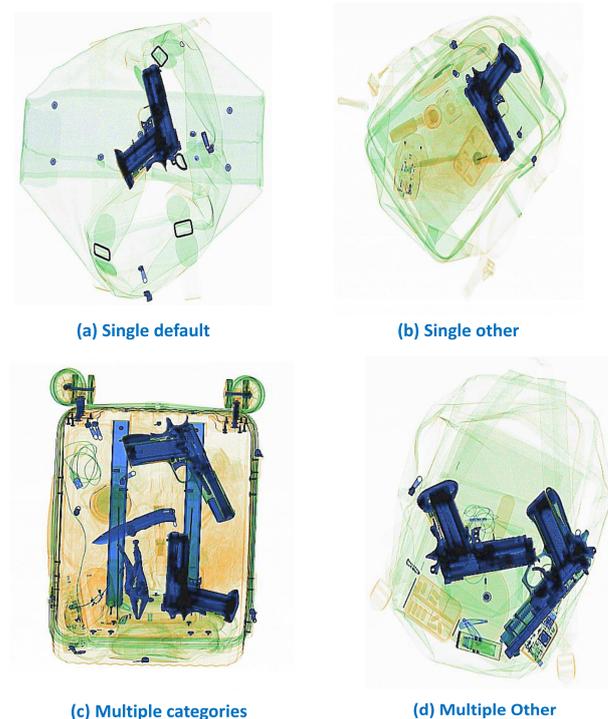


Figure 1. Examples of images from (a) Single Default, (b) Single Other, (c) Multiple Categories, and (d) Multiple Others file categories.

This dataset already included detailed bounding boxes and segmentation data in 140,000 images, offering two levels of precision to specify their precise location and shape. These segmentation data were essential for our next step: cropping the objects from the image by extracting the objects from their backgrounds. Figure 2 showcases the real images extracted from the original dataset and the different edges used in our experiment. This cropping process resulted in a collection of approximately 270,000 individual images.

Notably, the cropped dataset offered a high level of repetition, with numerous images showing the same or similar objects. To address this, we reduced our final dataset to 53,000 images by selecting only 20% of them.

We implemented this strategy to enhance computing speed and ensure diverse and representative samples for our study. It also allowed the model to create new datasets not present in the training sets.

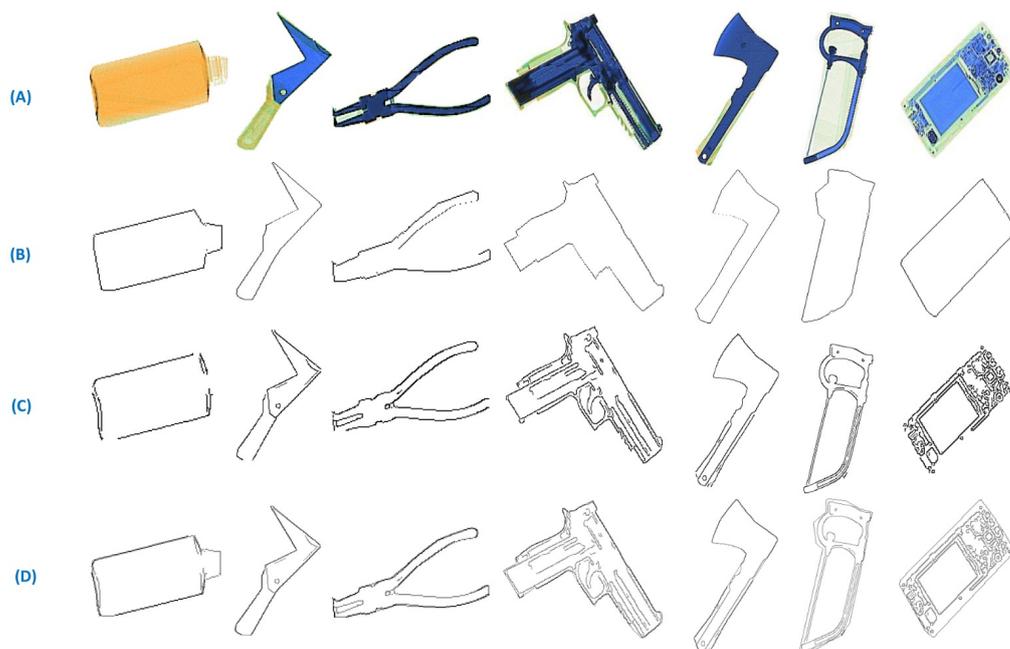


Figure 2. Examples of images used in our experiment (A) real images, (B) edges from mask, (C) edges from automated method, and (D) hybrid edges.

4. Methodology

To address the challenge of generating high-quality X-ray images of prohibited items, we explored image augmentation using generative adversarial networks (GANs) with Pix2Pix and Pix2PixHD as our core frameworks. We implemented a novel three-channel input technique to augment the model's ability to produce high-quality X-ray images. The first channel was dedicated to masks, assigning specific pixel values to distinctly identify object categories. The second and third channels incorporated edge data: Edge01, extracted from the shape of the object, and Edge02, generated through automated edge detection techniques on ground-truth images. The integration of Edge1 and Edge2, known as "Hybrid Edge", was designed to overcome the limitations of single-channel inputs, which often lacked detail and texture.

The combination of mask data with hybrid edges significantly improved the model's capability to generate images that were both visually accurate and detailed. The mask channel enabled precise object recognition and categorization, while the inclusion of both manually annotated and computationally derived edges allowed the model to depict a broad range of details, from sharp object outlines to subtle textural nuances essential for realistic X-ray imaging. Regarding the edge types employed:

Edge01: the shapes of objects were directly extracted from X-ray images, with a focus on capturing geometric outlines based on the annotations provided in the dataset [45].

Edge02: this involved converting images to greyscale, applying Gaussian blur to reduce noise, and using the Canny edge detection algorithm to enhance the detection of finer details and textures.

Hybrid Edge: A method that combined Edge01 and Edge02 to enhance image input data, improving both texture detail and shape accuracy. This approach overcame the limitations of Edge01's focus on outlines and Edge02's focus on internal details, providing a comprehensive solution for generating detailed and accurate images. This approach is illustrated in Figure 2.

These techniques collectively enable the model to simulate realistic and detailed X-ray images of prohibited items, highlighting the advantages of integrating diverse technological approaches in image synthesis. To achieve this high level of detail and realism, our methodology primarily utilised the Pix2Pix and Pix2PixHD models, which were pivotal in addressing the high resolution and detail-oriented requirements of our research. We chose

Pix2Pix and Pix2PixHD because of their exceptional capabilities in generating detailed, realistic images from paired datasets, whereas other models like CycleGAN, while adept at unpaired image translation, lacked the precision needed for X-ray security imaging due to their insufficient detail preservation [36].

Similarly, super-resolution techniques like SRGAN, although useful for enhancing resolution, do not facilitate the essential semantic-to-photo translation required. StyleGAN models, focusing primarily on artistic effects, did not meet the accurate semantic correspondences crucial for security applications. Also, DCGAN, known for its efficiency in generating quality images from noise, was unsuitable due to its inability to perform paired image translation [37], a key requirement for achieving the high fidelity necessary in our work. In contrast, Pix2Pix and Pix2PixHD, with their robust methodologies for handling paired data, provided a superior solution that aligned perfectly with the stringent needs of creating high-quality X-ray security images, ensuring both detail accuracy and high resolution.

Pix2Pix model: Introduced by P Isola [46] in 2016, it is a significant image-to-image translation framework using a conditional generative adversarial network (cGAN) architecture designed to learn mappings from input to output images using paired data sets from two distinct domains. The generator in Pix2Pix utilises a U-Net architecture, known for its efficient feature extraction and preservation through direct skip connections between encoder and decoder layers, enhancing stability in learning. The discriminator, called PatchGAN, evaluates patches of the image rather than the entire image, focusing on high-frequency details to ensure local realism with lower computational demands. The Figure 3 illustrates the architecture of the Pix2Pix model and the loss function of Pix2Pix can be expressed as:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \quad (1)$$

where x represents the input image, y is the corresponding real image in the target domain, and z is a random noise vector. This function is optimized in such a manner that G attempts to minimize this objective against an adversarial D that tries to maximize it, hence the term adversarial training.

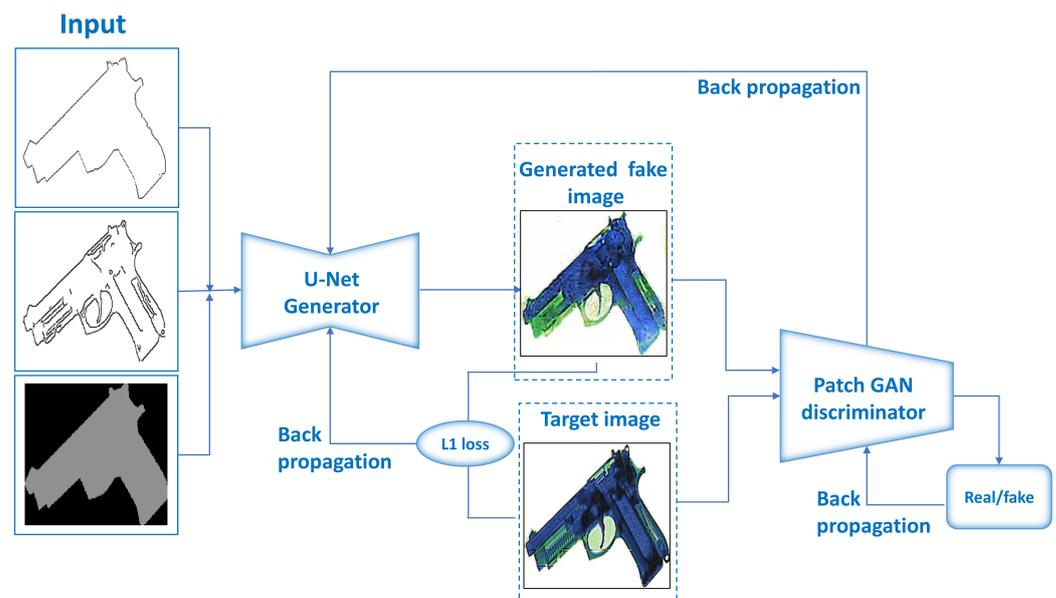


Figure 3. Pix2Pix conditional GAN model for image-to-image translation.

Pix2PixHD model: Developed by TC Wang [47], it enhances the original Pix2Pix framework by focusing on generating high-resolution images. This model is good at producing more detailed images through the use of multi-scale generator and discriminator

components. Pix2PixHD features a coarse-to-fine generator that includes a global generator network, which concentrates on broader and more general features like external contours and geometric structures, alongside a local enhancer network that focuses specifically on the finer details, such as the textures of objects. The discriminator network's architecture in pix2pixHD is similar to that of the original Pix2Pix, but it employs multiple discriminators, specifically, two in this case, that operate at different scales. This dual-discriminator setup enhances the generator's ability to create finely detailed X-ray images, as illustrated in Figure 4. The objective function of Pix2PixHD introduces an additional feature matching loss to ensure that the generated images not only deceive the discriminator but also closely align with the feature statistics of real images. The Pix2PixHD loss function is defined as:

$$\mathcal{L}_{FM}(G, D) = \mathbb{E}_{x,y} \sum_{i=1}^N \frac{1}{N_i} \left\| D^{(i)}(y) - D^{(i)}(G(x)) \right\|_1 \quad (2)$$

where $D^{(i)}$ represents the i th layer of the discriminator D , and N_i denotes the number of elements in each feature map $D^{(i)}$. This loss is added to the traditional adversarial loss, enhancing the model's ability to generate images that are visually indistinguishable from real images.

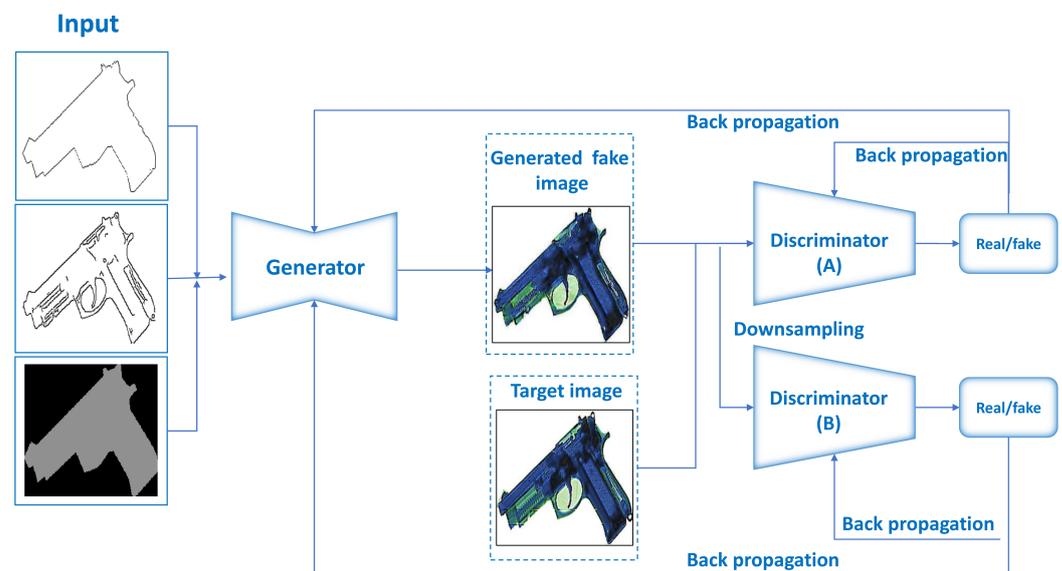


Figure 4. Pix2PixHD GAN model for image-to-image translation.

5. Experiments and Results

5.1. Experiment Setup

In our study, we applied Pix2Pix and Pix2PixHD models to a dataset of 53,000 X-ray image pairs across 35 object categories. The training consisted of 200 epochs, with a learning rate of 0.002 and a batch size of 128. We used an input image size of 256×256 pixels, as well as PyTorch libraries. Our desktop setup included an Intel i9-10900X processor, 192 GB of RAM, and four NVIDIA GeForce RTX4090 24 GB GPUs for efficient training.

5.2. Evaluation Metrics

To quantitatively assess the quality and realism of the generated images, we employed a comprehensive set of metrics, each chosen for its relevance to image synthesis and its acceptance in the research community for evaluating image quality:

5.2.1. PSNR (Peak Signal-to-Noise Ratio)

A measure of the quality of a reconstructed image compared to the original, expressed in decibels. Higher values indicate better quality.

$$\text{PSNR} = 20 \cdot \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \quad (3)$$

where

- \log_{10} is the base-10 logarithm;
- MAX_I is the maximum possible pixel value of the image;
- MSE represents the mean squared error between the original and the corrupted image, calculated without applying the square root.

5.2.2. SSIM (Structural Similarity Index Measure)

SSIM is a metric that evaluates the visual impact of three characteristics of an image: luminance, contrast, and structure, comparing changes between a reference and a processed image. Values range from -1 to 1 , where 1 indicates perfect similarity.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4)$$

where

- μ_x and μ_y are the average pixel values of images x and y , respectively;
- σ_x^2 and σ_y^2 are the variances of images x and y , respectively;
- σ_{xy} is the covariance between images x and y ;
- $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ are constants used to stabilize the division with a weak denominator. Here, L is the dynamic range of the pixel values (255 for 8-bit greyscale images), and $k_1 = 0.01$ and $k_2 = 0.03$ are default values.

5.2.3. Average LPIPS (Learned Perceptual Image Patch Similarity)

The average LPIPS is a deep learning-based metric that assesses the perceptual similarity between two images, focusing on human-like visual perception rather than pixel-level differences.

$$\text{LPIPS}(x, y) = \frac{1}{N} \sum_{i=1}^N d(\phi_i(x), \phi_i(y)) \quad (5)$$

where

- N is the number of image patches;
- $\phi_i(x), \phi_i(y)$ are the feature representations of the i th patch of images x and y , respectively;
- d is a distance function (e.g., Euclidean distance) that measures the difference between the feature representations.

5.2.4. FID Score (Fréchet Inception Distance)

The FID score compares the distribution of generated images to real images in the feature space of a deep neural network, measuring the quality and variety of generated images. Lower scores indicate a better quality and similarity to the real images.

$$\text{FID} = \|\mu_x - \mu_g\|^2 + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{1/2}) \quad (6)$$

- μ_x and μ_g are the featurewise means of the real and generated images, respectively.
- Σ_x and Σ_g are the covariance matrices of the real and generated images, respectively.
- Tr denotes the trace of a matrix, which is the sum of the elements on its main diagonal.

5.2.5. FSIM Score (Feature Similarity Index Matrix)

Using two primary attributes, Phase Congruency (PC) and gradient magnitude (GM), to assess image quality, the FSIM evaluates images based on these characteristics, encapsulating their assessment in the formula:

$$SL(x) = [SPC(x)]^\alpha \cdot [SG(x)]^\beta$$

Here, *PC* represents phase compatibility, crucial for identifying invariant features across different lighting conditions, while *GM* denotes the gradient magnitude, which highlights edge details by quantifying the rate of intensity change across the image. The parameters α and β allow the adjustment of the relative importance of these attributes in the final similarity score.

The gradient magnitude $GM(x)$ is calculated as follows:

$$GM(x) = \sqrt{G_x^2 + G_y^2}$$

where G_x and G_y are the horizontal and vertical gradients of the image, respectively.

The Phase Congruency (SPC) is computed with:

$$SPC = \frac{2 \cdot PC1 \cdot PC2 + T_1}{PC1^2 + PC2^2 + T_1}$$

In this formula, T_1 is a constant added to enhance the stability of the calculation, ensuring that *SPC* values range between 0 and 1. This approach effectively captures both the structural integrity and the perceptual likeness of the images, making the FSIM a robust tool for evaluating image quality in various applications.

5.3. Display of Results and Analysis Using Pix2Pix and Pix2PixHD with Masks

Pix2Pix with Masks: In this study, we prepared datasets to address the challenge of using masks as a single-channel input for generating new X-ray images. Our approach assigned a specific pixel value to each mask, uniquely representing different object categories. This method is particularly advantageous for X-ray imaging, where differentiating between elements requires a nuanced distinction. Each category within our dataset [45] was allocated a unique pixel value, facilitating a clearer differentiation among categories. An example of these masks is illustrated in Figure 5. To accommodate this labelling strategy, the Pix2Pix model was adapted; its generator output layer was modified to produce images where each pixel corresponded to the mask values, allowing the model to not only generate visual representations of objects but also classify them implicitly based on the mask's pixel values. Additionally, the discriminator was fine-tuned to assess both the authenticity of the generated images and the accuracy of object representation as per the mask labels, enhancing the model's ability to effectively distinguish between categories.



Figure 5. Examples of mask images for different object categories.

Pix2PixHD with masks: Following the Pix2Pix experiment, we applied the same dataset and masking technique as shown in Figure 5 to explore the performance of Pix2PixHD, a model designed for high-resolution image synthesis. We also used masks. This approach let us directly compare how well Pix2PixHD and the Pix2Pix model worked at generating new X-ray images. Given the advanced architecture of Pix2PixHD, which is capable of handling greater detail and producing higher-fidelity images, we anticipated an

improvement in the model's ability to generate visually accurate representations of objects, classify them based on the pixel values of the mask, and generate correct images.

The experiments conducted with the Pix2Pix and Pix2PixHD models, utilising masks to categorize objects in X-ray images, yielded significant findings. The visual results from the Pix2Pix model, as shown in Figure 6, did not meet the desired standards in terms of image quality and realism. Despite capturing the essence of the object categories represented by the masks, the generated images fell short of expectations with a PSNR of 11.50, SSIM of 0.41, LPIPS of 0.50, FID score of 166.44, and an FSIM score of 0.35. These results indicated the potential for improvement in enhancing the fidelity and realism of the generated images.

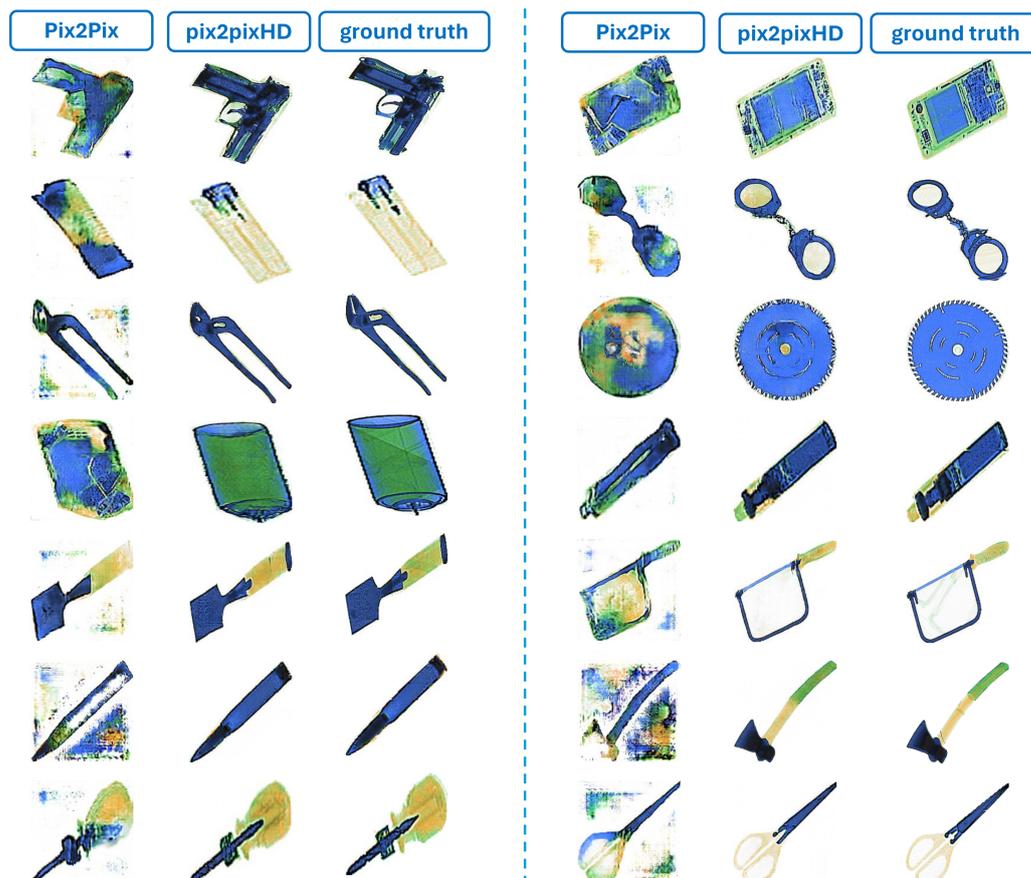


Figure 6. Detailed comparison of image synthesis using single-channel mask inputs across models—column 1: Pix2Pix output, column 2: Pix2PixHD output, column 3: corresponding ground-truth images.

In response to these findings, we conducted another experiment using Pix2PixHD to explore alternative approaches for generating more detailed and high-quality images. The Pix2PixHD model demonstrated markedly improved performance, achieving a PSNR of 16.43, SSIM of 0.60, LPIPS of 0.19, FID score of 35.36, and an FSIM of 0.46. This essential enhancement in the metric scores reflected Pix2PixHD's advanced capabilities in handling high-resolution and detailed image synthesis. The lower FID and high FSIM scores suggested a closer similarity of the generated images to the real X-ray images, demonstrating Pix2PixHD's superior ability to produce detailed and visually accurate representations of the objects.

Moreover, the visual results, as shown in Figure 6, closely approached the ground truth, confirming that Pix2PixHD was able to generate images remarkably similar to real X-ray images. The significant difference in FID and FSIM scores between Pix2Pix and

Pix2PixHD highlights the importance of choosing the right model for the image synthesis task, particularly when high fidelity is crucial for creating detailed outputs from masks as inputs.

This realisation motivated us to further explore Pix2PixHD’s capabilities by expanding our experiments to incorporate a variety of inputs. We conducted a number of additional experiments under the assumption that Pix2PixHD’s sophisticated architecture could more effectively exploit various types of input data to improve the detail of images. A detailed X-ray image is characterized by its exceptional precision in showing the smallest details. This level of detail allows the image to clearly display individual components, enabling the differentiation of various materials, the identification of objects, and the detection of anomalies with remarkable accuracy. Such detailed images provide a clearer view of the object’s shape, structure, and composition, which is important for accurately analysing and making correct security decisions. We aimed to explore the model’s adaptability and performance across a broader spectrum of input configurations, thereby deepening our understanding of its potential for complex image-generation tasks.

5.4. Detailed Experimentation and Results Analysis with Pix2PixHD

Given the exceptional performance Pix2PixHD displayed when given mask inputs, it became necessary to explore additional input variations in order to learn more about the model’s capabilities. Therefore, we conducted several experiments using various data inputs to fully understand Pix2PixHD’s ability to enhance X-ray image generation. These inputs comprised different combinations of edges and masks, allowing us to explore the model’s potential comprehensively.

5.4.1. Pix2PixHD with Different Types of Edge Inputs

In this experiment, we used edges instead of masks as inputs to our model. We utilised three distinct types of edges to conduct a comprehensive analysis. The effectiveness of different edge inputs for our model was quantitatively evaluated using a comprehensive set of metrics: the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS), Fréchet Inception Distance (FID), and Feature Similarity Index Metric (FSIM) scores. These scores provided clear evidence of the varied performance across the methods implemented, as summarized in Table 2.

Table 2. Quantitative impact analysis: effects of diverse inputs on synthetic X-ray image generation with Pix2Pix and Pix2PixHD models.

	PSNR	SSIM	LPIPS	FID Score	FSIM Score
Edge01	15.76	0.48	0.40	123.45	0.39
Edge02	17.62	0.56	0.31	78.48	0.46
Hybrid Edge	17.53	0.57	0.31	71.83	0.47
Mask–Pix2Pix	11.50	0.41	0.50	166.44	0.35
Mask–Pix2PixHD	16.43	0.60	0.19	35.36	0.46
Pix2Pix–edge and mask	13.89	0.54	0.35	146.20	0.42
Pix2PixHD–edge and mask	19.93	0.71	0.12	29.36	0.54

Specifically, the method utilising edges derived from X-ray images, denoted as Edge01, achieved a PSNR of 15.76, SSIM of 0.48, LPIPS of 0.40, FID score of 123.45, and FSIM score of 0.39 reflecting its capability but also highlighting areas for improvement. In contrast, Edge02 resulted in a higher PSNR of 17.62, SSIM of 0.56, LPIPS of 0.31, and a lower FID score of 78.48, and an FSIM score of 0.46, indicating better performance in image synthesis quality. Significantly, the combined approach, labelled Hybrid Edge, not only achieved an improved FID score of 71.837 and FSIM score of 0.47 but also showcased a PSNR of

17.53 and an SSIM of 0.57, with an LPIPS of 0.31. This progress highlighted the potential of combining various edge approaches to improve the accuracy of X-ray image generation.

Figure 7 presents a visual comparison of the same object-generated image using the three distinct edge input methods. The first image, generated with Edge01, illustrates that while the model was able to capture the overall shape of the object effectively, it fell short in rendering the complex details accurately. This is an indication of the method's ability to grasp general forms, but it has limitations in conveying the finer aspects of texture and small details within the X-ray images.

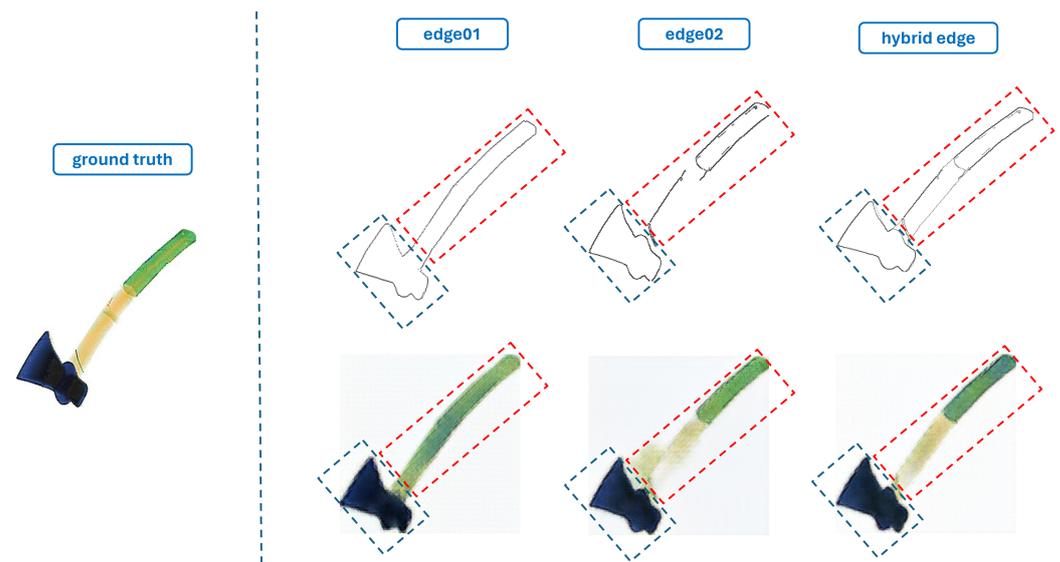


Figure 7. Analysis of different edge inputs and their synthesis results. Left to right: ground-truth images, edge inputs (Edge01, Edge02, Hybrid Edge) on the top line, and corresponding synthesis outcomes on the bottom line.

With Edge01, the results indicate that while the shape of objects was well-represented, the details within these objects were not captured effectively. This limitation arises because Edge01 focuses on the outline of objects, omitting finer details and textures inside the object's boundary. For example, in the case of the axe, the input successfully defined the axe's shape, but it did not capture the detailed distinction between different materials (wooden handle versus plastic component) in the handle of the axe. Also, we can notice that the head of the axe is not perfect and a little blurry, as shown in Edge01 in Figure 7. These results are due to the lack of internal detail information in the input, which causes Edge01 to not provide the details of the texture of the object like Edge02 and Hybrid Edge do.

Edge02 is designed to capture more of the internal details of objects compared to Edge01 by focusing on the texture and minor contrasts within the X-ray images, as well as the different object components.

As a result, Edge02 provided a significant improvement in generating the internal details of the axe's handle, differentiating areas of coloured green and orange (wooden handle versus plastic component) in the handle, with a good generation of the axe's head compared to Edge01. This suggests that Edge02 is effective in highlighting variations within objects that are not apparent with just the external contour information provided by Edge01. However, the technique also has limitations, as seen in the incomplete generation of the axe's handle's entirety. This incomplete part could result from the inherent challenges in the edge detection process, where certain features may not exhibit enough contrast after the greyscale conversion and Gaussian blur application in the input edge, making it difficult for the Canny algorithm to identify all relevant boundaries. From this experiment, we conclude that Edge02 can enhance the model's ability to generate internal object details,

leading to more detailed and textured representations, but it may also omit crucial parts of an object if the edge extraction is not complete in some cases. This partial representation can change the overall accuracy and realism of the images that are created, because important parts of objects might be misrepresented or completely missed. This can make the model less useful in situations where precise and detailed object reconstructions are needed.

Addressing the limitations found in generating X-ray images with Edge01 and Edge02, we used Hybrid Edge as the input, which is designed to provide a comprehensive solution that not only accurately delineates the object's shape but also captures its internal nuances and material differences with greater precision, as shown in Figure 8. The development of Hybrid Edge significantly improved object representation, as demonstrated by the method achieving the best scores on different edge inputs, confirming its superior effectiveness in generating high-quality and detailed X-ray images.

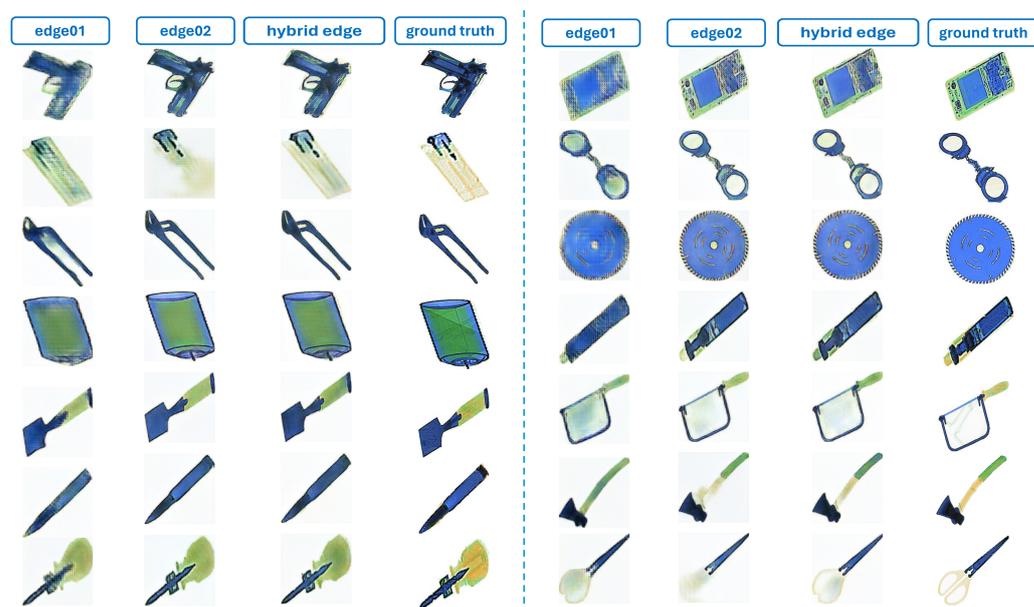


Figure 8. Comparative analysis of image synthesis with different edge inputs using the Pix2PixHD Model—column 1: Edge01 results, column 2: Edge02 results, column 3: Hybrid Edge results, and column 4: corresponding ground-truth image.

5.4.2. Pix2PixHD with Hybrid Edge and Mask Inputs

This experiment extended the methodology of the first two experiments by integrating mask inputs with both edge techniques to investigate their combined effect on the Pix2PixHD model's performance in generating high-quality X-ray images. Each input to the Pix2PixHD model consisted of three channels: one for the mask data, which assigned a specific pixel value to identify the object's category, and two for edge data, Edge01 and Edge02 (collectively termed "Hybrid Edge"), to emphasize textural details and subtle features often missed in the mask-only approach. The hybrid input was processed by the Pix2PixHD model to evaluate its effect on the quality of the generated images, as shown in Figure 9.

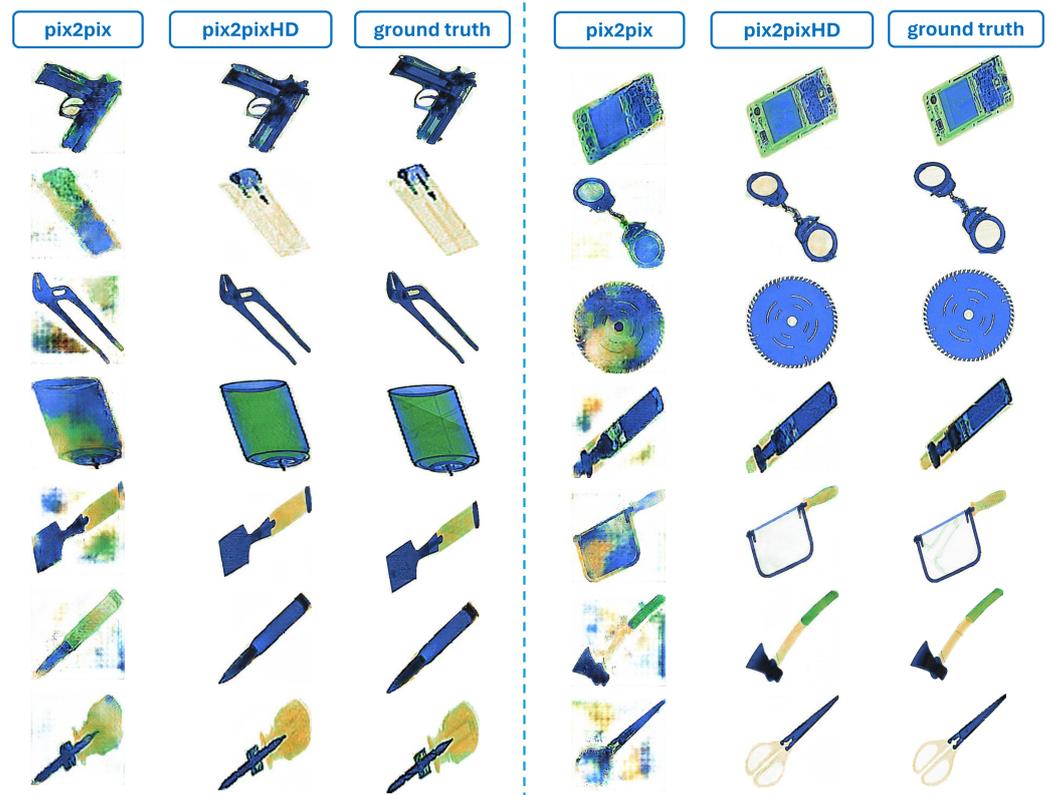


Figure 9. Comparative analysis when utilising masks and hybrid edges as three-channel inputs for image synthesis—column 1: Pix2Pix results, column 2: Pix2PixHD results, column 3: corresponding ground-truth images.

Our evaluation focused on comparing the results with those obtained from using only masks, paying particular attention to improvements in image detail, texture, and overall realism.

The inclusion of this enriched input strategy aimed to enhance Pix2PixHD's capabilities and compare its performance with the Pix2Pix model. This approach sets a new benchmark for image synthesis quality and enables further studies in the area of paired X-ray image security translation. Utilising the hybrid edge and mask input, the Pix2PixHD model achieved a PSNR of 19.93, SSIM of 0.71, LPIPS of 0.12, FID score of 29.36, and FSIM score of 0.54, exhibiting notably sharper and more detailed visual results compared to those from previous experiments with edge data, due to the comprehensive input details provided through the three-channel approach. Conversely, applying the same methodology to the Pix2Pix model yielded some improvements, with a PSNR of 13.89, SSIM of 0.54, LPIPS of 0.35, FID score of 146.20, and FSIM score of 0.42. However, the results confirmed that utilising a mask with hybrid edge inputs in conjunction with the Pix2PixHD model delivered the best outcomes, establishing a superior method for generating detailed and accurate X-ray images. Figure 10 shows the effectiveness of our proposed three-channel input.

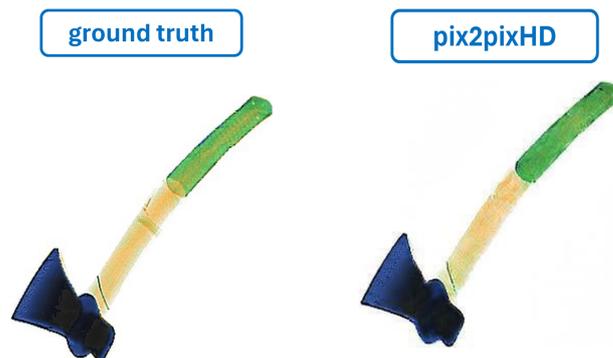


Figure 10. Result of utilising mask and hybrid edge as three-channel inputs with the axe from (Figure 9).

5.5. Comparative Analysis

Our comparative analysis of FID scores distinctly highlighted the superiority of our method over current state-of-the-art approaches in the field, as summarized in Table 3. Our technique achieved a remarkably low FID score of 29.36, significantly better than the scores reported in other notable studies. For instance, Jinfeng Yang [35] reported an FID score of 111.33 with only three categories, and Jk Dumagpi [2] achieved a score of 95.20 across five categories. Yue Zhu had two studies listed; one achieved an FID score of 69.50 with two categories [34], and another had an improved FID score of 37.04 with nine categories [37]. Additionally, D Liu’s work [36] also reported an FID score of 69.50 for two categories. The notable reduction in our FID score, despite the increased complexity of handling 35 categories, underscores our method’s effectiveness in producing high-quality images that more closely resemble real X-ray images. This result not only validates our innovative approach but also establishes a new benchmark for fidelity in image synthesis, facilitating the progression of future developments in this field.

Table 3. Comparative analysis of FID scores between our method and current state-of-the-art methods.

	FID	Number of Categories
Jinfeng Yang [35]	111.33	3
Jk Dumagpi [2]	95.20	5
Yue Zhu [34]	69.50	2
Y Zhu [37]	37.04	9
D Liu [36]	69.50	2
Our method	29.36	35

6. Conclusions and Future Work

In conclusion, our study showcased the remarkable capabilities of advanced generative models, particularly, Pix2Pix and Pix2PixHD, in synthesising realistic X-ray images for security screening applications. By augmenting the dataset with innovative techniques such as mask and hybrid edge inputs, we significantly improved image synthesis fidelity and accuracy. Our experiments conducted on a comprehensive Korean dataset demonstrated the superior performance of our novel mask and hybrid inputs, especially when using the Pix2PixHD model. The empirical evaluation confirmed the effectiveness of these generative models in overcoming the challenges of data scarcity in X-ray image synthesis. The metrics of PSNR, SSIM, LPIPS, and FID scores that we used in our quantitative analysis attested to the superiority of our proposed techniques. Furthermore, our research stood out by achieving the best FID score compared to other studies in the field, highlighting the quality and realism of the synthesised X-ray images. These findings confirm the potential of generative models for advancing X-ray screening technologies and open avenues for future research to further refine and expand these systems. In future studies, we aim

to convert camera images into X-ray images using advanced deep learning techniques. This approach will enable the generation of synthetic X-ray images from available camera images, expanding the variety and volume of training data for security screening models. Such advancements can significantly enhance the accuracy and robustness of detection systems. Additionally, we will focus on creating paired datasets comprising real and synthetic X-ray images, which are crucial for training and validating sophisticated deep learning models. These datasets will be pivotal in various fields, including security, medical imaging, and industrial inspection, aiming to train models that are more accurate and efficient. Moreover, a significant aspect of our future research will focus on enhancing the process of data layer separation in multi-object X-ray images within the context of X-ray security. By leveraging the high-quality synthetic X-ray images and their corresponding ground truths generated through our advanced modelling techniques, we plan to develop innovative methods for effectively distinguishing and isolating overlapping objects within complex scans. This capability will be particularly useful when creating complex images featuring multiple objects in X-ray security settings, each with its own isolated ground truth. We also aim to expand the creation of complex multi-object datasets for X-ray security, which will enhance the application and utility of X-ray images.

Author Contributions: Conceptualization, B.Y. and M.S.K.; Methodology, B.Y. and M.S.K.; Software, B.Y.; Validation, B.Y. and M.S.K.; Formal analysis, B.Y. and H.-S.K.; Resources, H.-S.K.; Data curation, B.Y.; Writing – original draft, B.Y.; Writing – review & editing, B.Y. and M.S.K.; Visualization, B.Y.; Supervision, H.-S.K.; Project administration, H.-S.K.; Funding acquisition, H.-S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Research Projects of “Development of automatic screening and hybrid detection system for hazardous material detecting in port container” funded by the Ministry of Oceans and Fisheries (20200611, 70%), and partly by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (IITP-2024-2020-0-01462, 30%).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gaus, Y.F.A.; Bhowmik, N.; Akcay, S.; Breckon, T. Evaluating the Transferability and Adversarial Discrimination of Convolutional Neural Networks for Threat Object Detection and Classification within X-Ray Security Imagery. In Proceedings of the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 420–425.
2. Dumagpi, J.K.; Jeong, Y.J. Evaluating gan-based image augmentation for threat detection in large-scale xray security images. *Appl. Sci.* **2020**, *11*, 36. [[CrossRef](#)]
3. Han, L.; Ma, C.; Liu, Y.; Jia, J.; Sun, J. SC-YOLOv8: A Security Check Model for the Inspection of Prohibited Items in X-ray Images. *Electronics* **2023**, *12*, 4208. [[CrossRef](#)]
4. Juneja, M.; Minhas, J.S.; Singla, N.; Kaur, R.; Jindal, P. Denoising techniques for cephalometric x-ray images: A comprehensive review. In *Multimedia Tools and Applications*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 1–39.
5. Du, Y.B.; Jia, R.; Cui, Z.; Yu, J.T.; Sun, H.M.; Zheng, Y. X-ray image super-resolution reconstruction based on a multiple distillation feedback network. *Appl. Intell.* **2021**, *51*, 5081–5094. [[CrossRef](#)]
6. Basiricò, L.; Ciavatti, A.; Fraboni, B. Solution-Grown Organic and Perovskite X-Ray Detectors: A New Paradigm for the Direct Detection of Ionizing Radiation. *Adv. Mater. Technol.* **2021**, *6*, 2000475. [[CrossRef](#)]
7. Bhowmik, N.; Breckon, T.P. Joint Sub-component Level Segmentation and Classification for Anomaly Detection within Dual-Energy X-Ray Security Imagery. In Proceedings of the 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas, 12–14 December 2022; pp. 1463–1467.
8. HR, S.K.; Bhargavi, M.; Kumar, C., P. Classification of COVID-19 and Pneumonia X-ray Images Using a Transfer Learning Approach. In Proceedings of the 2021 IEEE Region 10 Symposium (TENSymp), Grand Hyatt Jeju, Republic of Korea, 23–25 August 2021; pp. 1–6.

9. Li, S.; Liu, W.; Xiao, G. Detection of Srew Nut Images Based on Deep Transfer Learning Network. In Proceedings of the 2019 Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 951–955.
10. Masita, K.L.; Hasan, A.N.; Paul, S. Pedestrian detection using R-CNN object detector. In Proceedings of the 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Guadalajara, Mexico, 7–9 November 2018; pp. 1–6.
11. Mahmoud, M.; Kang, H.S. GANMasker: A Two-Stage Generative Adversarial Network for High-Quality Face Mask Removal. *Sensors* **2023**, *23*, 7094. [[CrossRef](#)]
12. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-form image inpainting with gated convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 4471–4480.
13. Hu, Z.; Tang, J.; Wang, Z.; Zhang, K.; Zhang, L.; Sun, Q. Deep learning for image-based cancer detection and diagnosis- A survey. *Pattern Recognit.* **2018**, *83*, 134–149. [[CrossRef](#)]
14. Yagoub, B.; Ibrahim, H.; Salem, A.; Kang, H.S. Single energy x-ray image colorization using convolutional neural network for material discrimination. *Electronics* **2022**, *11*, 4101. [[CrossRef](#)]
15. Minaee, S.; Liu, Z. Automatic question-answering using a deep similarity neural network. In Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, Canada, 14–16 November 2017; pp. 923–927.
16. Abdallah, A.; Kasem, M.; Hamada, M.A.; Sdeek, S. Automated Question-Answer Medical Model based on Deep Learning Technology. In Proceedings of the 6th International Conference on Engineering & MIS 2020, Larnaka, Cyprus, 9–11 June 2020; pp. 1–8.
17. Alsayat, A. Customer decision-making analysis based on big social data using machine learning: A case study of hotels in Mecca. *Neural Comput. Appl.* **2023**, *35*, 4701–4722. [[CrossRef](#)]
18. Kasem, M.S.; Hamada, M.; Taj-Eddin, I. Customer profiling, segmentation, and sales prediction using AI in direct marketing. *Neural Comput. Appl.* **2024**, *36*, 4995–5005. [[CrossRef](#)]
19. Mahmoud, M.; Kasem, M.; Abdallah, A.; Kang, H.S. AE-LSTM: Autoencoder with LSTM-Based Intrusion Detection in IoT. In Proceedings of the 2022 International Telecommunications Conference (ITC-Egypt), Alexandria, Egypt, 26–28 July 2022; pp. 1–6.
20. Xu, W.; Jang-Jaccard, J.; Singh, A.; Wei, Y.; Sabrina, F. Improving performance of autoencoder-based network anomaly detection on nsl-kdd dataset. *IEEE Access* **2021**, *9*, 140136–140146. [[CrossRef](#)]
21. Bakator, M.; Radosav, D. Deep learning and medical diagnosis: A review of literature. *Multimodal Technol. Interact.* **2018**, *2*, 47. [[CrossRef](#)]
22. Lata, K.; Cenkeramaddi, L.R. Deep learning for medical image cryptography: A comprehensive review. *Appl. Sci.* **2023**, *13*, 8295. [[CrossRef](#)]
23. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
24. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE Computer Society: Washington, DC, USA, 2017; pp. 6517–6525.
25. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.E.; Fu, C.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*; Springer International Publishing: Cham, Switzerland, 2015; pp. 21–37.
26. Fang, C.; Liu, J.; Han, P.; Chen, M.; Liao, D. FSVM: A Few-Shot Threat Detection Method for X-ray Security Images. *Sensors* **2023**, *23*, 4069. [[CrossRef](#)] [[PubMed](#)]
27. Yu, X.; Yuan, W.; Wang, A. X-ray Security Inspection Image Dangerous Goods Detection Algorithm Based on Improved YOLOv4. *Electronics* **2023**, *12*, 2644. [[CrossRef](#)]
28. Gao, Q.; Deng, H.; Zhang, G. A Contraband Detection Scheme in X-ray Security Images Based on Improved YOLOv8s Network Model. *Sensors* **2024**, *24*, 1158. [[CrossRef](#)]
29. Brock, A.; Donahue, J.; Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. *arXiv* **2018**, arXiv:1809.11096.
30. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [[CrossRef](#)]
31. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
32. Yagoub, B.; Ibrahim, H.; Salem, A.; Suh, J.W.; Kang, H.S. X-ray image denoising for cargo dual energy inspection system. In Proceedings of the 2021 International Conference on Electronics, Information, and Communication (ICEIC), Jeju, Republic of Korea, 31 January–3 February 2021; pp. 1–4.
33. Akcay, S.; Breckon, T. Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging. *Pattern Recognit.* **2022**, *122*, 108245. [[CrossRef](#)]
34. Zhu, Y.; Zhang, Y.; Zhang, H.; Yang, J.; Zhao, Z. Data augmentation of X-ray images in baggage inspection based on generative adversarial networks. *IEEE Access* **2020**, *8*, 86536–86544. [[CrossRef](#)]
35. Yang, J.; Zhao, Z.; Zhang, H.; Shi, Y. Data augmentation for X-ray prohibited item images using generative adversarial networks. *IEEE Access* **2019**, *7*, 28894–28902. [[CrossRef](#)]

36. Liu, D.; Liu, J.; Yuan, P.; Yu, F. A data augmentation method for prohibited item X-ray pseudocolor images in X-ray security inspection based on Wasserstein generative adversarial network and spatial-and-channel attention block. *Comput. Intell. Neurosci.* **2022**, *2022*, 8172466. [[CrossRef](#)]
37. Zhu, Y.; Zhang, H.G.; An, J.Y.; Yang, J.F. GAN-based data augmentation of prohibited item X-ray images in security inspection. *Optoelectron. Lett.* **2020**, *16*, 225–229. [[CrossRef](#)]
38. Zhao, Z.; Zhang, H.; Yang, J. A GAN-based image generation method for X-ray security prohibited items. In Proceedings of the Pattern Recognition and Computer Vision: First Chinese Conference, PRCV 2018, Guangzhou, China, 23–26 November 2018; Proceedings, Part I 1; Springer: Berlin/Heidelberg, Germany, 2018; pp. 420–430.
39. Shao, F.; Liu, J.; Wu, P.; Yang, Z.; Wu, Z. Exploiting foreground and background separation for prohibited item detection in overlapping X-Ray images. *Pattern Recognit.* **2022**, *122*, 108261. [[CrossRef](#)]
40. Qiu, L.; Xiong, Z.; Wang, X.; Liu, K.; Li, Y.; Chen, G.; Han, X.; Cui, S. ETHSeg: An Amodel Instance Segmentation Network and a Real-world Dataset for X-Ray Waste Inspection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2283–2292.
41. Dumagpi, J.K.; Jeong, Y.J. Pixel-level analysis for enhancing threat detection in large-scale X-ray security images. *Appl. Sci.* **2021**, *11*, 10261. [[CrossRef](#)]
42. Dumagpi, J.K.; Jeong, Y.J. End-to-End Object Separation for Threat Detection in Large-Scale X-Ray Security Images. *IEICE Trans. Inf. Syst.* **2022**, *105*, 1807–1811. [[CrossRef](#)]
43. Mery, D.; Rizzo, V.; Zscherpel, U.; Mondragón, G.; Lillo, I.; Zuccar, I.; Lobel, H.; Carrasco, M. GDxray: The database of X-ray images for nondestructive testing. *J. Nondestruct. Eval.* **2015**, *34*, 42. [[CrossRef](#)]
44. Miao, C.; Xie, L.; Wan, F.; Su, C.; Liu, H.; Jiao, J.; Ye, Q. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2119–2128.
45. X-ray Images of Hazardous Items. Available online: <https://aihub.or.kr/aidata/33> (accessed on 24 October 2023).
46. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
47. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional GANs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8798–8807.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.