FGL-GAN: Global-Local Mask Generative Adversarial Network for Flame Image Composition

Kui Qin, Xinguo Hou *, Zhengjun Yan, Feng Zhou and Leping Bu

School of Electrical Engineering, Naval University of Engineering, Wuhan 430033, China
* Correspondence: hxinguo2008@126.com

Abstract: It is important to reduce the danger of collecting flame image data sets by compositing flame images by computer. In this paper, a Global-Local mask Generative Adversarial Network (FGL-GAN) is proposed to address the current status of low quality composite flame images. First, FGL-GAN adopts a hierarchical Global-Local generator structure, to locally render high-quality flame halo and reflection, while also maintaining a consistent global style. Second, FGL-GAN incorporates the fire mask as part of the input of the generation module, which improves the rendering quality of flame halo and reflection. A new data augmentation technique for flame image compositing is used in the network training process to reconstruct the background and reduce the influence of distractors on the network. Finally, FGL-GAN introduces the idea of contrastive learning to speed up network fitting and reduce blurriness in composite images. Comparative experiments show that the images composited by FGL-GAN have achieved better performance in qualitative and quantitative evaluation than mainstream GAN. Ablation study shows the effectiveness of the hierarchical Global-Local generator structure, fire mask, data augmentation, and MONCE loss of FGL-GAN. Therefore, a large number of new flame images can be composited by FGL-GAN, which can provide extensive test data for fire detection equipment, based on deep learning algorithms.

Keywords: composite flame image; generative adversarial networks; Global-Local; fire mask; data augmentation

1. Introduction

In recent years, with the rapid development of artificial intelligence, fire detection algorithms [1,2] based on computer vision have been widely used. However, due to safety concerns, ignition experiments cannot be performed in many indoor environments, such as warehouses, hangars, oil depots, etc. This makes it impossible for fire detection algorithms to test fire images in these scenarios. If the flame (foreground) collected in laboratories can be combined with the above-mentioned indoor environments (background), the above issue can be well addressed.

The simplest image compositing method is to extract the flame from the real image and paste it into the background (cut-paste algorithm called for short, as shown in Figure 1). If the flame image composited by the cut-paste algorithm is compared with the real image, it is evident that there is a marked difference between the composited image and the real image. This is mainly due to the following two reasons. First, there is a halo around the real flame, therefore the flame can transition smoothly into the background and there is no obvious boundary. The cut-paste algorithm cannot paste the halo around the flame in the new background, which renders the composite flame different in appearance from the real flame. Second, the real flame itself is a light source, which will cast reflection on the ground. However, the flame composited by the cut-paste algorithm cannot produce reflection, which makes the composite image unrealistic.
Researchers have made considerable effort in order to address the above problems, which is mainly divided into two categories: traditional image processing and deep learning.

Among traditional methods, the most popular method is to force gradient domain smoothing of the foreground [3–5], the most classic of which is Poisson image editing [6]. It maintains the gradient domain between the foreground and the background by computing the gradient of the background and passing it on to the foreground. Traditional methods change the flame’s color and texture information, based on the background information. However, the flame is used as the light source, whose color and texture are not affected by the background during the compositing process. Therefore, traditional methods are not suitable for solving the problem of flame image compositing.

In deep learning methods, image compositing can be considered as an image translation problem, in which the cut-paste composite image is mapped to the real image. However, image translation often makes random changes to the style and content of the background or objects. Therefore, if the image translation algorithm is simply applied to flame image compositing, it can easily cause flame or background distortion.

At present, there is little research on the problem of flame image compositing. Yang et al. proposed FIS-GAN [7]. This network has two independent GANs. The first GAN controls the texture of the flame when the flame is generated by adding code to the input noise. The second GAN reduces blurring of the composite image by adding an attention mechanism to the generator in flame image compositing. However, there are still three problems with this method. First, FIS-GAN is mainly aimed at 128 × 128 image compositing with low resolution and little background information, which will lead to an unsatisfactory fusion effect between background and foreground. Second, FIS-GAN can only generate flame reflection with a fixed area. Under normal circumstances, the area of flame reflection changes with the color, shape of the flame, and background light intensity. The fixed area flame reflection limits the use of FIS-GAN. Finally, the realism of the flame halo and reflection generated by FIS-GAN needs to be improved.

In order to overcome the shortcomings of the above methods, we propose a novel flame image compositing algorithm: FGL-GAN. The network aims to achieve realistic flame image compositing. FGL-GAN uses generative adversarial ideas to learn from the real flame images to composite new flame images. The main contributions of this paper can be summarized into the following four aspects:

1. For the problem of flame image compositing, we propose a new generative adversarial network ‘FGL-GAN’. The generator structure of FGL-GAN is divided into two parts: local generation and global coordination. The local generation part is responsible for generating flame halo and reflection. The global coordination part is responsible for fusing the generated local image with the background. The two parts cooperate with each other to jointly optimize the quality of the composite image;

2. To improve the performance of FGL-GAN, the fire mask and a new data augmentation are proposed in this paper. The fire mask extracted from the original image becomes part of the input to the generation module in FGL-GAN, which can pass information
such as the position and shape of the flame to the network, thereby improving the quality of the rendering of flame halo and reflection. Some images of non-flame scenes are also used as training sets for FGL-GAN, to achieve reconstruction of different scene backgrounds and reduce the influence of distractors. At the same time, FGL-GAN introduces the idea of contrastive learning and uses the Modulated Noise Contrastive Estimation (MONCE) loss [8] as part of the overall loss;

3. Extensive comparative experiments and ablation study demonstrate that FGL-GAN achieves the best results for flame image compositing. FGL-GAN has better effectiveness and wider applicability than other networks, and can provide lots of flame datasets for flame detection.

2. Related Works

This paper mainly incorporates the method of image translation to study the problem of image compositing. Therefore, this section mainly introduces the research status of image translation and image compositing.

2.1. Image Translation

GAN proposed in [9–11] makes pixel-level image translation [12–14] possible, and has been widely used in image style transfer [15,16], image segmentation [19,20], super-resolution [21,22] and other fields. GAN consists of a generator and a discriminator. The generator generates a fake image, and the discriminator is used to distinguish the generated image from the real image. The generated image is constantly approaching the real image through the adversarial learning between them.

Later, most of the image translation networks adopted the framework of generative adversarial networks. Isola et al. proposed pix2pix [23], which for the first time used conditional generative adversarial networks as a general solution for supervised image translation. But the network’s datasets must be matched. CycleGAN [24] achieved image translation for non-matching datasets by combining G:X (source domain) → Y (target domain) and F:Y → X mappings, and introducing a cycle consistency loss function. However, the network has two mappings, resulting in increasing the amount of network parameters and training time. The above two networks can only achieve one-to-one image translation.

In order to achieve translation between multiple datasets, starGAN [25] uses different mask vectors for different datasets. HistoGAN [26] achieves controllable image translation by encoding the color histogram of the input image and feeding it into the network. To improve the resolution of the translated image, Ting-Chun et al. proposed multi-scale generator and discriminator [27], which achieves 2048 × 1024 resolution of the translated image. The translated images of pix2pix and cycleGAN are relatively blurry. Park et al. believed that the general normalization layer loses part of the image information; as a result, they proposed a spatially-adaptive normalization layer [28] to generate realistic images. Hu et al. designed a QS-atten module [29] to improve the quality of the generated images, without changing the model parameters.

However, the above methods basically deal with image translation of public datasets, and do not perform well in solving the problem of flame image compositing. Therefore, this paper proposes FGL-GAN for flame image compositing.

2.2. Image Compositing

Image compositing [30–32] includes, but is not limited to, the following four issues: object placement, image blending, image harmonization, and shadow generation. Object placement aims to paste the foreground into the background at a reasonable position and size. Refs. [33,34] tries to find a suitable position for the foreground via learning the prior knowledge of the background. Refs. [35,36] predicts the position for the foreground via learning the relationship between the foreground and the background. Image blending aims to refine the foreground boundary and reduce blurriness. The simplest approach is alpha blending [37]. Ref. [38] achieves a smooth transition on the boundary by using the gradient
domain constraint as a loss function, according to the Poisson equation, Ref. [39] blends the foreground and background into a seamless new image by using two independent encoders to extract multi-scale characteristics from both. Image harmonization aims to unify the color and lighting characteristics of the foreground and background. Refs. [40,41] count the appearance features of foreground and background (such as color histogram) to unify them through linear or nonlinear transformation. Ref. [42] unifies their appearance by learning the appearance features of foreground and background through an additional spatial separation attention module. Ref. [43] makes the foreground and background consistent by learning the geometry, color, and boundary features between them. Shadow generation aims to generate shadows for foreground objects based on the lighting information of the background, to make the composite image look more realistic. Ref. [44] proposes a lighting inference algorithm that can recover the complete lighting information of the scene from a single image. Ref. [45] designs an interactive software which can generate corresponding soft shadows under user control, according to the lighting information of the background.

Of the four steps of flame image compositing, since manual placement of flame position is more reasonable compared with machine placement, manual placement is chosen as the object placement in this paper. The flame itself is a light source and does not need to be harmonized by the background. Therefore, the flame image compositing can be simplified into two steps: image blending and shadow generation. Image blending is mainly the blurring transition of the boundary between the flame and the background, and the generation of halo around the flame. Shadow generation is mainly the generation of flame reflection. These two issues can be addressed with the image translation model. Therefore, this paper builds an image translation model, FGL-GAN, to solve these two problems together.

3. The Proposed Model: FGL-GAN

This section introduces the FGL-GAN proposed in this paper in detail, in four parts: network structure, mask, data augmentation, and loss.

3.1. Network Structure

FGL-GAN generates realistic flame halo and reflection through adversarial learning between generator and discriminator. The generator structure of the network is shown in Figure 2. The discriminator structure of the network is shown in Figure 3. The generator and discriminator are introduced in detail below.

3.1.1. Generator

The generator is a Global-Local structure, whose learning goal is to composite realistic flame images. It contains two generation modules: local generation module (LGM) and global coordination module (GCM). LGM aims to make the network focus on local information (around the flame), thereby generating clearer, more realistic, and more detailed flame halo, and reflection. However, if the generator only contains LGM, the color of the generated local image and the background will be inconsistent, and the boundary between the local image and the background will be too sharp, resulting in an unnatural and unrealistic overall image. Therefore, the generator blends the local image with the background color and blurs the boundaries through GCM, so that the final composite image is more natural and realistic. The two generation modules are coordinated and optimized together to composite the final flame image.

The generator contains two local operation modules: local region section module (LRSM) and local region recovery module (LRRM). The generator selects the local region to be composited through the LRSM as the input of the LGM. The generator maps the local image back to the overall image through LRRM as input to GCM. The generator uses LRSM and LRRM to separate flame halo and reflection generation and global coordination into two parts, thus reducing the difficulty of each part.
3.1.1. Generator

The generator is a Global-Local structure, whose learning goal is to composite realistic flame images. It contains two generation modules: local generation module (LGM) and global coordination module (GCM). LGM aims to make the network focus on local information (around the flame), thereby generating clearer, more realistic, and more detailed flame halo, and reflection. However, if the generator only contains LGM, the color of the generated local image and the background will be inconsistent, and the boundary between the local image and the background will be too sharp, resulting in an unnatural and unrealistic overall image. Therefore, the generator blends the local image with the

The generator also contains a segmentation module (SM) designed to implement flame image segmentation. SM uses the segmented local fire mask ($M_{\text{fire},l}$) and global fire mask ($M_{\text{fire},g}$) as the input of LGM and GCM, respectively. The two fire masks enable LGM and GCM to focus on flame and its surroundings, instead of image style transfer, thereby improving the quality of composite images.

The steps of image compositing inside the generator are: $I_i$ is obtained through cut-paste algorithm (as shown in Figure 1). The generator receives $I_i$ and local mask $M_l$ obtained by the cut-paste algorithm. First, the local image $L_i$ is selected from $I_i$ by LRSM. The local fire mask $M_{\text{fire},l}$ is segmented from $L_i$ by SM. Then $L_i$ combined with $M_{\text{fire},l}$ are
inputted into LGM to render flame halo and reflection. LGM outputs local composite image \( L_g \). \( L_g \), \( I_i \), and \( M_L \) are composited into global image \( G_g \) by LRRM. Finally, global fire mask \( M_{\text{fire,}g} \) is obtained by \( I_i \) through SM. \( M_{\text{fire,}g} \) combined with \( G_g \) and are inputted into GCM for global coordination. GCM outputs the final composite flame image \( I_c \).

3.1.2. Discriminator

The goal of the discriminator is to distinguish the authenticity of the composite image. It contains two sub discrimination modules: local discriminator module (\( D_L \)) and global discriminator module (\( D_G \)). When \( D_L \) has \( L_g \) as input, the output is probability that \( L_g \) is false. When \( D_L \) has \( L_R \) as input, the output is probability that \( L_R \) is real. In order to make the output \( L_g \) of LGM consistent with the input \( L_i \) in texture, \( L_g \) and \( L_i \) or \( L_R \) and \( L_i \) are combined as the input to \( D_L \). \( D_G \) is similar to \( D_L \), except that the inputs are replaced with global \( I_c \) and \( G_R \).

3.2. Mask

Compared with the general image translation network, in order to make the generator focus on changing the perimeter of the flame rather than the overall style of the image, this paper designed two masks in FGL-GAN: fire mask and local mask.

Fire mask contains \( M_{\text{fire,}l} \) and \( M_{\text{fire,}g} \). They are obtained by segmenting the flame from local or global images, and are used as part of the inputs to the two generation modules. Because the generation modules do not want to change the color and texture of the flame, the generator inputs the fire mask to the generation module and directly conveys the position, shape, size, and other information of the flame, to the generation module. This makes the generation module focus on the flame halo and reflection generation. Fire mask can avoid the occurrence of flame shape deformation, color distortion, and the change of overall style of the image in the process of flame image compositing.

Local mask (\( M_L \)) is artificially set. LRSM takes \( M_L \) as the criterion to select a specific local region \( L_i \) from the initial \( I_i \) according to Equation (1). LRRM takes \( M_L \) as the criterion, and combines the flame local image \( L_g \) generated by LGM, and the initial \( I_i \), into a new global image \( G_g \) according to Equation (2). \( M_L \) is the valve of LRSM and LRRM. Users can control \( M_L \) to change the flame position and region of halo and reflection. The size of the local mask should not be fixed. The size of the local mask is artificially set and can affect the area of the flame reflection.

\[
L_i(x, y) = \begin{cases} 
I_i(x, y) & M_L(x, y) = 1 \\
\text{NaN} & M_L(x, y) = 0 
\end{cases}
\quad (1)
\]

where \((x, y)\) is the pixel of the image.

\[
G_g(x, y) = \begin{cases} 
L_g(x, y) & M_L(x, y) = 1 \\
I_i(x, y) & M_L(x, y) = 0 
\end{cases}
\quad (2)
\]

3.3. Data Augmentation

The reasons why data augmentation technology is used in this paper are as follows: First, the flame datasets used in this paper have only 24 different scenes. When there are fewer scenes, the network is prone to overfitting, resulting in blurred or even distorted background of the composite images. Second, the network should learn to recognize flames. It can render flame halo and reflection in scenes with fire, and only completes the background reconstruction in flameless scenes. However, when there are fewer scenes, it is difficult for the network to complete the background reconstruction in flameless scenes. Finally, when there are fewer training scenes, the network is easily affected by distractors (objects similar to flames, such as strong lights, car lights, flashlights), and is prone to generate halo and reflection around the distractors. Therefore, data augmentation is necessary.
In order to avoid the above problems, this paper uses a large number of flameless scenes as part of the network training set. The purpose is so that the network not only learns the compositing of flame images in scenes with flame, but also learns the reconstruction of the background in the flameless scenes (as shown in Figure 4, the size and location of the local mask is a random setting when the input image is a flameless image). The two goals are not in conflict because, in the fire scenes, except in regions around the fire, most of the other background regions should also be reconstructed. Therefore, the data augmentation in this paper is to input the flameless scenes into the network with a certain probability, as shown in the following equation:

$$I_t = \alpha I_{fire} + (1 - \alpha)I_{no\_fire}$$  \hspace{1cm} (3)

where $\alpha$ is the proportional coefficient for adjusting the flame scenes and the flameless scenes.

3.4. Loss

In order to better solve the problem of flame image compositing, this paper designed four sub losses, including: local adversarial loss, local L1 loss [23], global adversarial loss, and MONCE loss [8].

3.4.1. Local Adversarial

Adversarial loss is common in GANs [46–48]. Its purpose is to make the generator and discriminator learn adversarially. The local discriminator strives to distinguish the local image, $L_R$, generated by the LGM in the generator, from the real local image, $L_R$, but the goal of LGM is to render flame reflection and halo as realistic as possible. Local adversarial loss is defined as:

$$L^L_{GAN}(G, D_L) = E_{L_R,i}\left[\log D_L(L_i, L_R)\right] + E_L[1 - \log D_L(L_i, L_R)]$$  \hspace{1cm} (4)
3.4.2. Local L1 Loss

In order to ensure the similarity between the generated local image, $L_g$, and the real local image, $L_R$, this paper uses L1 regularization to calculate the error between the generated image and the real image, pixel by pixel. Local L1 loss is defined as:

$$L_{L1}^L(G) = \| L_R - L_g \|_1$$

(5)

3.4.3. Global Adversarial Loss

The global discriminator strives to distinguish global image, $I_c$, composited by the generator, from the real global image, $G_R$, while the goal of the generator is to make the composite flame images as consistent as possible with the real flame images. Global adversarial loss is defined as:

$$L_{GAN}^G(G, D_L) = E_{I_i, G_R}[ \log D_G(I_i, G_R)] + E_{I_i}[1 - \log D_G(I_i, I_c)]$$

(6)

3.4.4. MONCE loss

In order to reduce the blurriness of composite images, this paper introduces contrastive losses [49–51]. However, the general contrastive losses tend to set all negative samples to the same weight, which will reduce the performance of the generator. However, MONCE designates a negative sample similarity target matrix to retrieve an optimal negative sample weight through optimal transport [52], which can improve the quality of the composite images. This paper adopts the MONCE loss and it is defined as:

$$L_{MONCE}^N(X, Y) = -\sum_{i=1}^{N} \log \frac{e^{x_i \cdot y_i / \tau}}{e^{x_i \cdot y_i / \tau} + \sum_{j=1}^{N} w_{ij} \cdot e^{x_i \cdot y_i / \beta}}$$

(7)

where $X = [x_1, x_2, \ldots, x_N]$ are image feature sets encoded from composite images $I_c$. $Y = [y_1, y_2, \ldots, y_N]$ are image feature sets encoded from real images $G_R$. $\tau$ is the temperature parameter. $N$ is the number of feature patches. $Q$ is the weight of negative terms. $w_{ij}$ is the coefficient obtained by optimal transport (as shown in Equation (8)).

$$\min_{w_{ij}, i, j} \left[ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \cdot e^{x_i \cdot y_i / \beta} \right] \text{ s.t. } \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} = 1, \sum_{i=1}^{N} w_{ij} = 1$$

(8)

where $\beta$ is the temperature parameter of the cost matrix in optimal transport.

The overall loss is the sum of the above losses and is defined as:

$$L_{total}(G, D_L, D_G) = L_{GAN}^G + \lambda_1 L_{GAN}^G + \lambda_2 L_{L1}^L + \lambda_3 L_{MONCE}$$

(9)

where $\lambda_1$, $\lambda_2$, $\lambda_3$ are the weights for adjusting each loss.

4. Implementation

This section introduces the implementation details of the network from the following three aspects: datasets production, generator structure, and discriminator structure.

4.1. Datasets Production

At present, there is no common datasets for flame image composite. To produce the datasets, we took 24 fire videos of different scenes through experiments, converted them into 5000 images, and divided them into training sets, validation sets, and test sets by 8:1:1 ratio. Furthermore, we searched the Internet for some images of warehouses and hangars (without fire) as part of the training sets.
The production process of the datasets is shown in Figure 5. Details are as follows: First, local mask is obtained by artificially marking the flame and the area of its halo, and reflection. Second, the flame is segmented from the original image. Then, the flame is pasted directly to the background image according to Equation (10). Finally, the flame, flame halo, and reflection, are labeled from the original image to form a local mask $M_L$.

$$I_i = G_s \cdot M_{fire,g} + B \cdot [1 - M_{fire,g}]$$  \hspace{1cm} (10)

where $G_s$ is flame which is segmented from the original image. $B$ is the background. $M_{fire,g}$ is the fire mask. $I_i$ is the image where the flame is pasted directly to the background.

4.2. Generator Structure

LRSM and LRRM have been introduced above and their details will not be repeated here. This subsection mainly introduces the two generation modules and the segmentation modules.

4.2.1. Generation Modules

The two generation modules are the classic Encoder-Resnet-Decoder structure (as shown in Figure 6). Encoder-Decoder is a symmetrical structure, which ensures the consistency of input and output image sizes. Resnet network [53] introduces residual error, which improves the learning ability of the network. The two generation modules are only different in the number of layers of the resnet block. Due to the small size of local images, the LGM has seven layers, but the GGM has 10 layers. The detailed structure of each block in the generator is Appendix A.

Figure 5. Datasets production process.
4.2.2. Segmentation Modules

With reference to [54,55], the obtained flame images have the following unique color characteristics, compared with images from other objects: (1). R component value of the flame is higher than the other component values. (2). The B component value and B component dispersion of the flame and the distractors are quite different. (3). The S component of the flame is linearly related to the R component. Using the above color characteristics of the flame images, the criterion for the flame segmentation module designed in this paper is shown in Equation (11). Equation (11) can retain flame information, while suppressing interferences such as flame reflection and sunlight, with good segmentation effect.

\[
f(x, y) = \begin{cases} 
  f(x, y) & \text{if } \begin{array}{l}
    R \geq G \geq B \& B \geq B_T \\
    \& S > (1 - R)S_T / R_T \& B_{\text{std}} \geq B_{\text{std}, T}
  \end{array} \\
  0 & \text{else}
\end{cases}
\]

where \( f(x, y) \) is the color value at each position in the image \( R, G, B, S \) are the red, green, blue, and saturation components of the image, respectively. \( B_T, R_T, S_T \) are blue component threshold, red component threshold, and saturation component threshold, respectively. \( B_{\text{std}} \) is the blue component dispersion of a certain area in the image [55]. \( B_{\text{std}, T} \) is blue component dispersion threshold.

4.3. Discriminator Modules

The local discriminator module adopts the structure of PatchGAN (as shown in Figure 7). PatchGAN has no limit on the size of input images and outputs an \( M \times N \) matrix. Each element in the matrix represents the authenticity of a patch in the original image, therefore the final target value is the average of all elements in the matrix. The global discriminator adopts the structure of DCGAN [56] and outputs the probability of the authenticity of the global image (The detailed structure of the global discriminator is Appendix A).
5. Experiment Results

The FGL-GAN designed in this paper, along with current mainstream image translation networks including pix2pix [23], cycleGAN [24], QS-attn [29], and FIS-GAN [7], which is the only generative adversarial network for flame image compositing, was iterated 400,000 times under the same conditions on RTX2080Ti. The results are evaluated and compared qualitatively and quantitatively.

5.1. Quantitative Evaluation Metrics

The quantitative evaluation in this paper is gauged in three aspects: the comparison between composite images and real images, computer vision, and user evaluation.

*FID* is selected for the comparison between composite images and real images. *FID* refers to the process in which the composite images are inputted into the trained inception [57] network, and Fréchet Inception Distance [58] is calculated with features extracted from the layer before the last fully connected layer. *FID* can evaluate the similarity between composite images and real images, and the clarity and diversity of composite images. The lower the *FID* value, the better the effect of image compositing.

Resnet accuracy (acc) [53] and yolov5 [59] confidence(conf) were selected for computer vision. Resnet accuracy refers to the classification accuracy of fire and non-fire on the composite images, from the same trained resnet network. The higher the classification accuracy, the better the authenticity of the composite images. Yolov5 confidence refers to the probability that a certain object in the composite image is determined to be fire, from a trained yolov5 network. The higher the yolov5 confidence, the more realistic the composite image, and the better the rendering of flame halo and reflection.

User evaluation refers to the process in which 10 users select one with the best composite effect, from a group of images composited by all networks. The score for one network is the ratio of the best-performing images composited by that network, to the total number of composite images. The evaluation process is divided into global and local. From the global perspective, users mainly compare the authenticity of the image composition. From the local perspective, users mainly compare the generation effects of flame halo and reflection.

5.2. Qualitative Evaluation

The qualitative evaluation in this paper was carried out on test sets and prediction sets. The real images of the test sets were known, the background and flame position were the same as the training sets, but the flame was different to the training sets. The real images of the prediction sets were unknown, and the background, flame, and flame position, were all different to the training sets.

One example of the test sets is shown in Figure 8. In Figure 8, it can be observed that the reflective area rendered by the cycleGAN composite images is randomly larger, or smaller, than the real images. The flame halo and reflection rendered by the QS-attn composite images are relatively blurry, and are quite different from the real images. The composite images by FIS-GAN have a clear boundary between the flame reflection and the background, and is relatively blurry. Although the flame halo and reflection rendered by pix2pix and FGL-GAN are more reasonable, the halo and reflection rendered by FGL-GAN are closer to the real images, than pix2pix.

One example of the prediction sets is shown in Figure 9. The background light of the example images range from light to dark. In Figure 9, it can be observed that cycleGAN can composite more realistic images when the light is bright, but can only generate halo, and basically cannot render reasonable reflection when the light is moderate or dark. Pix2pix can composite more realistic images when the light is moderate. However, the images composited by pix2pix are distorted (the entire ground becomes a flame reflective area) when the light is bright. Additionally, pix2pix cannot render reasonable reflection when the light is dark. The overall image style of composite images rendered by QS-attn is easily altered, which makes the images unreal. Also, flame halo and reflection rendered by QS-attn are poor among all the networks. There is a clear boundary between local
image composited by FIS-GAN and the background, which makes the composite images look unreal. Flame halo and reflection rendered by FIS-GAN are unreasonable: flame halo rendered by FIS-GAN is poor, which makes the border between the flame and the background too sharp. Normal flame reflection should spread from the bottom of the flame towards the ground, and diminish gradually. However, the flame reflection rendered by FIS-GAN always clusters under the flame, and its aberration is not obvious.

![Figure 8. The example of the test sets.](image)

![Figure 9. An example of the prediction sets.](image)

From a global perspective, the boundary between flame and background in the images composited by FGL-GAN is smooth, its background style will not change, it will not distort, and it is more natural and realistic. From a local perspective, the flame halo and reflection rendered by FGL-GAN are more reasonable. First, the flame halo and reflection change with the background light. When the light is strong and natural light is the main light source, the reflection of the flame is not obvious; as the light weakens, the flame gradually becomes the main light source and the reflection becomes more and more obvious. Second, the rendered reflection of the flame gradually attenuates with increasing distance from the flame. Finally, the rendered flame halo and reflections are less affected by distractors. Compared with other networks, the FGL-GAN proposed in this paper has the best effect on composited flame images.
source, the reflection of the flame is not obvious; as the light weakens, the flame gradually becomes the main light source and the reflection becomes more and more obvious. Second, the rendered reflection of the flame gradually attenuates with increasing distance from the flame. Finally, the rendered flame halo and reflections are less affected by distractors. Compared with other networks, the FGL-GAN proposed in this paper has the best effect on composited flame images.

5.3. Quantitative Evaluation

This paper tested the FID, resnet accuracy, yolov5 confidence, and global and local user evaluation, composited by each network. The results are shown in Table 1. The FID comparison results show that compared with other networks, the FGL-GAN proposed in this paper has the lowest FID of 29.75, which indicates that the halo and reflection rendered by FGL-GAN are closest to the real images, and the clarity and diversity of the composite images are the best. From the perspective of computer vision algorithm evaluation, resnet accuracy, and the yolov5 confidence of FGL-GAN, achieve 0.9386 and 0.7534, respectively, which are both better than other networks. It shows that the flame halo and reflection composited by FGL-GAN can “deceive” the fire classification and recognition algorithms, based to certain extent on computer vision, and can achieve the purpose of “mixing the fake with the real”. From the perspective of user evaluation, compared with other networks, users incline to choose the images composited by FGL-GAN, whether for the authenticity of the image as a whole or for local rendering of the flame halo and reflection. This shows the images composited by FGL-GAN are more in line with the aesthetics of human vision, and are easier to deceive humans. To sum up, the quality of the flame images composited by FGL-GAN proposed in this paper excels, compared to the other methods.

|                      | cycleGAN | pix2pix | QS-attn | FIS-GAN | FGL-GAN |
|----------------------|----------|---------|---------|---------|---------|
| FID                  | 47.10    | 40.29   | 59.52   | 53.25   | 29.75   |
| Computer vision acc  | 0.7778   | 0.7222  | 0.7389  | 0.9115  | 0.9386  |
| conf                 | 0.6067   | 0.5788  | 0.5928  | 0.6828  | 0.7534  |
| User evaluation global | 0.167   | 0.092   | 0.125   | 0.033   | 0.583   |
| local                | 0.027    | 0.118   | 0.040   | 0.179   | 0.636   |

5.4. Ablation Study

We conducted ablation study to evaluate the effectiveness of each module of the method proposed in this paper, which can be divided into five cases: only LGM, only GCM, no MONCE loss, no fire mask, and FGL-GAN.

The example of test sets is shown in Figure 10, and the example of prediction sets is shown in Figure 11. In Figures 10 and 11, compared with only LGM and FGL-GAN, only LGM produced an obvious border between the local composite image and the background, which would cause the overall image to be unreal, while FGL-GAN adds GCM after LGM, which would make the local image and the background blend better and make the images more realistic. Compared with only GCM and FGL-GAN, the flame halo and reflection rendered by only GCM were more blurred than FGL-GAN, which shows that adding the LGM before the GCM can effectively improve the clarity of rendering flame halo and reflection in composite images. Compared with no MONCE loss or FGL-GAN, the composite images without the former loss in the prediction sets were obviously distorted, and there was an obvious boundary between the local image and the background. Therefore, adding the MONCE loss to the network can prevent the composite images from distortion. Compared with no fire mask and FGL-GAN, flame halo and reflection in the prediction sets were basically not rendered with no fire mask. FGL-GAN adds fire mask to the input of the two generation modules of the network, which can provide shape, position,
size of the flame, and other information to the network, thereby rendering reasonable flame halo and reflection.

Figure 10. Ablation study of test sets.

Figure 11. Ablation study of prediction sets.

The comparison of \( FID \) and yolov5 confidence in five cases is shown in Table 2. \( FID \) comparison results show: \( FID \) of FGL-GAN is slightly better than no fire mask and with only global coordination module, indicating that adding local generation module and fire mask to the network can improve resemblance between composite images and real images, as well as the clarity of composite images. There is a big gap in \( FID \) between FGL-GAN and the two cases where there was only LGM and no MONCE loss in the network, indicating that the GCM and MONCE loss in the network can significantly improve the quality of composite images. Yolov5 confidence comparison results show: FGL-GAN (combining with LGM and GCM) has better confidence than a single module, indicating effectiveness of FGL-GAN co-optimizing the two modules. Compared with no MONCE and no fire mask, the confidence of FGL-GAN can be significantly improved. The main reasons are shown in Figure 11: the network with no MONCE loss can cause distorted composite images in the prediction sets. The network with no fire mask can cause significantly reduced effect in halo and reflection rendering.

Table 2. Comparison of \( FID \) and yolov5 confidence in different situations of ablation study.

|                 | Only GCM | Only LGM | No MONCE | No Fire Mask | FGL-GAN |
|-----------------|----------|----------|----------|--------------|---------|
| \( FID \)       | 33.08    | 51.37    | 51.21    | 31.40        | 29.75   |
| \( \text{conf} \) | 0.7332   | 0.7046   | 0.6682   | 0.6915       | 0.7534  |
This paper explores the background reconstruction and the influence of distractors on the network before and after data augmentation, as shown in Figure 12. Compared with before and after data augmentation, the background reconstruction is relatively blurred, and the network is easily affected by interference, which can cause flame halo and reflection near alarm lights or headlights before data augmentation. However, the use of data augmentation can significantly improve the clarity of background reconstruction and reduce the impact of distractors on the network.

![Background Reconstruction](image1)

![Interference effects](image2)

**Figure 12.** Comparison with before and after, using data augmentation.

This paper explores the impact of using MONCE loss on network fitting speed, as shown in Figure 13. It can be observed that using MONCE loss can significantly speed up network fitting.

![Variation of FID](image3)

**Figure 13.** Variation of FID with the number of training epochs.
In addition, an experiment was carried out to study the effect of the size of the local mask on the area of flame reflection. Detailed experiments are in Appendix B.

6. Conclusions

This paper proposes a novel generative adversarial network, FGL-GAN, for flame image compositing. First, FGL-GAN aims to achieve local and global co-optimization through the Global-Local structure. Second, the fire mask is inputted to the generation module to improve the quality of flame halo and reflection rendering. Then, the use of new data augmentation techniques aims to enhance the immunity of the network to interference. Finally, FGL-GAN introduces MONCE loss to improve the network’s performance.

Qualitative comparative evaluations of FGL-GAN with existing methods show FGL-GAN can composite better flame images from the global and local perspective in different circumstances, compared to other methods. Quantitative comparative evaluations of FGL-GAN with existing methods show the images composited by FGL-GAN reach 29.75 on the FID, reach 0.9386 on resnet accuracy, reach 0.7534 on yolov5 confidence, reach 0.583 on global user evaluation, and reach 0.636 on local user evaluation. The results of FGL-GAN are all better than the other compared methods.

A series of ablation experiments demonstrated that the hierarchical Global-Local generator structure, fire mask, and the MONCE loss of FGL-GAN, can significantly improve the quality of composited images. The data augmentation of FGL-GAN improves the anti-interference performance of the network. The MONCE loss can significantly speed up network fitting.

In summary, relatively realistic virtual fire images can be composited by FGL-GAN in places where fire experiments cannot be performed. Therefore, the images composited by FGL-GAN can be used to test fire detection algorithms.

In Figure 9, it can be observed that the quality of FGL-GAN composite images when the light is moderate, is better than when light is too bright or too dark. This is mainly because the training sets are mostly images with moderate light. To solve this problem, we will continue to optimize FGL-GAN.

Author Contributions: Conceptualization, K.Q., X.H., Z.Y., F.Z. and L.B.; data curation K.Q., Z.Y. and F.Z.; methodology K.Q., X.H. and L.B.; software K.Q. and Z.Y.; validation F.Z.; formal analysis K.Q., Z.Y., F.Z. and L.B.; investigation K.Q. and X.H.; writing—original draft preparation K.Q., X.H. and L.B.; writing—review and editing K.Q., F.Z. and Z.Y.; supervision X.H. and L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Appendix A

The generator and discriminator structures in FGL-GAN are described in detail in this appendix:

The LGM and GGM in the generator takes the structure of “Encoder-Resnet-Decoder”. The specific structure of each block is shown in Figure A1.
Appendix B

An experiment was carried out to study the effect of the size of the local mask on the area of flame reflection. The input images are combined with different local mask in this experiment. They are all fed into FGL-GAN to composite images. The experimental results are shown in Figure A3. It can be observed that the area of flame reflection will increase with the size of the local mask.
Figure A3. The effect of the size of the local mask on the area of flame reflection.

References

1. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A review on early forest fire detection systems using optical remote sensing. *Sensors* **2020**, *20*, 6442. [CrossRef][PubMed]
2. Gaur, A.; Singh, A.; Kumar, A.; Kapoor, K. Video flame and smoke based fire detection algorithms: A literature review. *Fire Technol.* **2020**, *56*, 1943–1980. [CrossRef]
3. Fattal, R.; Lischinski, D.; Werman, M. Gradient domain high dynamic range compression. In Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, San Antonio, TX, USA, 23–26 July 2002; Volume 7, pp. 249–256.
4. Kazhdan, M.; Hoppe, H. Streaming multigrid for gradient-domain operations on large images. *ACM Transactions on graphics (TOG)*. **2008**, *27*, 1–10.
5. Levin, A.; Zomet, A.; Peleg, S.; Weiss, Y. Seamless image stitching in the gradient domain. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 377–389.
6. Pérez, P.; Gangnet, M.; Blake, A. Poisson image editing. *ACM SIGGRAPH 2003 Pap.* **2003**, 313–318. [CrossRef]
7. Yang, Z.; Wang, T.; Bu, L.; Jineng, O. Training with Augmented Data: GAN-based Flame-Burning Image Synthesis for Fire Segmentation in Warehouse. *Fire Technol.* **2022**, *58*, 183–215. [CrossRef]
8. Zhan, F.; Zhang, J.; Yu, Y.; Wu, R.; Lu, S. Modulated contrast for versatile image synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 21 June 2022; pp. 18280–18290.
9. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *3*, 2672–2680.
10. Karras, T.; Aittala, M.; Laine, S.; Hellsten, J.; Lehtinen, J.; Aila, T. Alias-free generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 852–863.
11. Lin, J.; Zhang, R.; Ganz, F.; Han, S.; Zhu, J. Anycost gans for interactive image synthesis and editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14986–14996.
12. Pizzati, F.; Cerri, P.; de Charette, R. CoMoGAN: Continuous model-guided image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14288–14298.
13. Pang, Y.; Lin, J.; Qin, T.; Chen, Z. Image-to-image translation: Methods and applications. *IEEE Trans. Multimed.* **2021**, *24*, 3859–3881. [CrossRef]
14. Ruan, C.; Yuan, L.; Hu, H.; Chen, D. Image translation with dual-directional generative adversarial networks. *IET Comput. Vis.* **2021**, *15*, 73–83. [CrossRef]
15. Lin, T.; Ma, Z.; Li, F.; He, D.; Li, X.; Ding, E.; Wang, N.; Li, J.; Gao, X. Drafting and revision: Laplacian pyramid network for fast high-quality artistic style transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 5141–5150.
16. Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; Aila, T. Training generative adversarial networks with limited data. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12104–12114.
17. Guo, X.; Yang, H.; Huang, D. Image Inpainting via Conditional Texture and Structure Dual Generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 14134–14143.
21. Zhang, K.; Liang, J.; Van Gool, L.; Timofte, R. Designing a practical degradation model for deep blind image super-resolution. arXiv 2021, 2103.14006. Available online: https://arxiv.org/abs/2103.14006 (accessed on 8 March 2022).

22. De Farias, E.C.; Di Noia, C.; Han, C.; Sala, E.; Castelli, M.; Rundo, L. Impact of GAN-based lesion-focused medical image super-resolution on the robustness of radiomic features. Sci. Rep. 2021, 11, 12361. [CrossRef]

23. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, 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.

24. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.

25. Choi, Y.; Choi, M.; Kim, M.; Ha, J.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.

26. Afifi, M.; Brubaker, M.A.; Brown, M.S. Histogan: Controlling colors of gan-generated and real images via color histograms. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7941–7950.

27. 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.

28. Park, T.; Liu, M.-Y.; Wang, T.-C.; Zhu, J.-Y. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2337–2346.

29. Hu, X.; Zhou, X.; Huang, Q.; Shi, Z.; Sun, L.; Li, Q. QS-Attn: Query-Selected Attention for Contrastive Learning in I2I Translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–20 June 2022; pp. 18291–18300.

30. Niu, L.; Cong, W.; Liu, L.; Hong, Y.; Zhang, B.; Liang, J.; Zhang, L. Making Images Real Again: A Comprehensive Survey on Deep Image Composition. arXiv 2021, 2106.14490. Available online: https://arxiv.org/abs/2106.14490 (accessed on 24 February 2022).

31. Zhan, F.; Lu, S.; Zhang, C.; Ma, F.; Xie, X. Adversarial image composition with auxiliary illumination. arXiv 2020, arXiv:2009.08255. Available online: https://arxiv.org/abs/2009.08255 (accessed on 25 February 2022).

32. Azadi, S.; Pathak, D.; Ebrahimi, S.; Darrell, T. Compositional GAN: Learning Image-Conditioned Binary Composition. arXiv 2018, arXiv:1807.07560. Available online: https://arxiv.org/abs/1807.07560 (accessed on 6 February 2022). [CrossRef]

33. Remez, T.; Huang, J.; Brown, M. Learning to segment via cut-and-paste. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 37–52.

34. Fang, H.-S.; Sun, J.; Wang, R.; Gou, M.; Li, Y.-L.; Lu, C. Instaboost: Boosting instance segmentation via probability map guided copy-pasting. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 682–691.

35. Tan, F.; Bernier, C.; Cohen, B.; Ordonez, V.; Barnes, C. Where and who? automatic semantic-aware person composition. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 15 March 2018; pp. 1519–1528.

36. Zhan, F.; Zhu, H.; Lu, S. Spatial fusion gan for image synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3653–3662.

37. Porter, T.; Duff, T. Compositing digital images. In Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques, Minneapolis, MN, USA, 23–27 July 1984; pp. 253–259.

38. Wu, H.; Zheng, S.; Zhang, J.; Huang, K. Gp-gan: Towards realistic high-resolution image blending. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2487–2495.

39. Zhang, H.; Zhang, J.; Perazzi, F.; Patel, V. Deep image composting. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2021; pp. 365–374.

40. Xue, S.; Agarwala, A.; Dorsey, J.; Rushmeier, H. Understanding and improving the realism of image composites. ACM Trans. Graph. (TOG) 2012, 31, 1–10. [CrossRef]

41. Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. IEEE Comput. Graph. Appl. 2001, 21, 34–41. [CrossRef]

42. Cun, X.; Pun, C.M. Improving the harmony of the composite image by spatial-separated attention module. IEEE Trans. Image Processing 2020, 29, 4759–4771. [CrossRef]

43. Chen, B.C.; Kae, A. Toward realistic image composting with adversarial learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8415–8424. [CrossRef]

44. Karsch, K.; Sunkavalli, K.; Hadap, S.; Carr, N.; Jin, H.; Fonte, R.; Sittig, M.; Forsyth, D. Automatic scene inference for 3d object composting. ACM Trans. Graph. (TOG) 2014, 33, 1–15. [CrossRef]
45. Sheng, Y.; Zhang, J.; Benes, B. SSN: Soft shadow network for image compositing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4380–4390.
46. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 7354–7363.
47. Brock, A.; Donahue, J.; Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. *arXiv* 2018, arXiv:1809.11096. Available online: https://arxiv.org/abs/1809.11096 (accessed on 15 March 2022).
48. Noguer, J.; Contreras, I.; Mujahid, O.; Aleix, B.; Josep, V. Generation of Individualized Synthetic Data for Augmentation of the Type 1 Diabetes Data Sets Using Deep Learning Models. *Sensors* 2022, 22, 4944. [CrossRef] [PubMed]
49. Ting, C.; Simon, K.; Mohammad, N.; Geoffrey, H. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, Virtual, 12–18 July 2020; pp. 1597–1607.
50. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, U.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised contrastive learning. *Adv. Neural Inf. Process. Syst.* 2020, 33, 18661–18673.
51. Kaiming, H.; Haoqi, F.; Yuxin, W.; Saining, X.; Ross, G. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9729–9738.
52. Peyré, G.; Cuturi, M. Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* 2019, 11, 355–607. [CrossRef]
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
54. Chen, T.-H.; Wu, P.-H.; Chio, Y.-C. An early fire-detection method based on image processing. In Proceedings of the 2004 International Conference on Image Processing, Singapore, 24–27 October 2004; pp. 1707–1710.
55. Wang, T.; Bu, L.; Yang, Z.; Yuan, P.; Ouyang, J. A new fire detection method using a multi-expert system based on color dispersion, similarity and centroid motion in indoor environment. *IEEE/CAA J. Autom. Sin.* 2019, 7, 263–275. [CrossRef]
56. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* 2015, arXiv:1511.06434. Available online: https://arxiv.org/abs/1511.06434 (accessed on 16 February 2022).
57. Szegedy, C.; Liu, W.; Jia, Y.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]
58. Dowson, D.C.; Landau, B.V. The Fréchet distance between multivariate normal distributions. *J. Multivar. Anal.* 1982, 12, 450–455. [CrossRef]
59. Jocher, G.; Stoken, A.; Borovec, J. UltraLyt-Ics/Yolov5: v 5.0-YOLOv5-P6 1280 Models. AWS, Su-pervisely and YouTube Integrations, Zenodo, 2021. Available online: https://github.com/ultralytics/yolov5 (accessed on 14 March 2022).