Article

Real-Time Video Smoke Detection Based on Deep Domain Adaptation for Injection Molding Machines

Ssu-Han Chen [1], Jer-Huan Jang [2,†], Meng-Jey Youh [2,†], Yen-Ting Chou [1,†], Chih-Hsiang Kang [3,†], Chang-Yen Wu [4], Chih-Ming Chen [4], Jiun-Shiung Lin [1], Jin-Kwan Lin [5] and Kevin Fong-Rey Liu [6,*]

[1] Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei City 243303, Taiwan; ssuhanchen@mail.mcut.edu.tw (S.-H.C.); hjhjkoko2022@gmail.com (Y.-T.C.); jslin@mail.mcut.edu.tw (J.-S.L.)

[2] Department of Mechanical Engineering, Ming Chi University of Technology, New Taipei City 243303, Taiwan; jhjang@mail.mcut.edu.tw (J.-H.J.); mjyouh@mail.mcut.edu.tw (M.-J.Y.)

[3] Center of Artificial Intelligent and Data Science, Ming Chi University of Technology, New Taipei City 243303, Taiwan; c.h.kang@o365.mcut.edu.tw

[4] 1st Petrochemicals Division, Formosa Chemicals & Fibre Corporation, Taipei City 105076, Taiwan; cywu.chem1@fcfc.com.tw (C.-Y.W.); cmc01.chem2@fcfc.com.tw (C.-M.C.)

[5] Department of Business and Management, Ming Chi University of Technology, New Taipei City 243303, Taiwan; jinkwan@mail.mcut.edu.tw

[6] Department of Safety, Health and Environmental Engineering, Ming Chi University of Technology, New Taipei City 243303, Taiwan

* Correspondence: kevinliu@mail.mcut.edu.tw

† These authors contributed equally to this work.

Abstract: Leakage with smoke is often accompanied by fire and explosion hazards. Detecting smoke helps gain time for crisis management. This study aims to address this issue by establishing a video smoke detection system, based on a convolutional neural network (CNN), with the help of smoke synthesis, auto-annotation, and an attention mechanism by fusing gray histogram image information. Additionally, the study incorporates the domain adversarial training of neural networks (DANN) to investigate the effect of domain shifts when adapting the smoke detection model from one injection molding machine to another on-site. It achieves the function of domain confusion without requiring labeling, as well as the automatic extraction of domain features and automatic adversarial training, using target domain data. Compared to deep domain confusion (DDC), naïve DANN, and the domain separation network (DSN), the proposed method achieves the highest accuracy rates of 93.17% and 91.35% in both scenarios. Furthermore, the experiment employs t-distributed stochastic neighbor embedding (t-SNE) to facilitate fast training and smoke detection between machines by leveraging domain adaption features.

1. Introduction

With the rapid industrial development and the emergence of Industry 4.0, manual labor has been gradually replaced by automated unmanned production lines, leading to increased production capacity. As a result, safety, health, and environmental (SHE) protections have become major concerns for enterprises. In recent years, smart factories have gained prominence, focusing on the efficient utilization of personnel, raw materials, equipment, and the environment. Enterprises have been implementing various measures, such as real-time monitoring, production automation, and information integration, by introducing factory production lines.

Injection molding, a widely used manufacturing method in the plastic injection molding industry, involves melting raw materials into a liquid state at high temperatures, giving them plastic properties. Smart factories' injection molding production lines typically consist of several computer numerical control (CNC) machines, skilled professionals, and smart systems. However, the high temperature associated with the injection molding process poses a fire safety issue. Therefore, managers emphasize fire prevention measures in the workshop to mitigate the risk of fire accidents.

It is important to note that smoke often accompanies fire or explosion incidents. Consequently, fire alarms incorporating smoke detection have been widely deployed in production lines. Leveraging the advancements in digital image processing (DIP) technology and computer vision (CV) applications, visual fire alarms have gained widespread implementation. Unlike traditional methods that rely on temperature and smoke particle detection, DIP can identify flames and smoke visually. However, the operational deployment of these systems in industrial settings presents several challenges, including issues with the light source, human movement, complex backgrounds, and variations in smoke patterns. Advancements in graphics processing units (GPUs), deep learning (DL) algorithms, and the reduced cost of computer computing resources have improved the field of visual fire alarm systems. These advancements ensure a safer working environment in smart factories and other production facilities.

Many researchers have developed new DIP smoke detection methods because smoke detection can be applied to different scenarios, and different DIP technologies have been developed to extract smoke features according to various conditions. Mutar and Dway proposed a dynamic smoke detection method [1]. This method detected smoke, based on frame movement, through the characteristics of early smoke. Smoke was detected with features extracted by binarizing each frame of the smoke image, removing pixels according to the brightness value, and using standard deviation to measure the gray level and transparency of the smoke image. Wu et al. used texture, wavelet, color, edge orientation histogram, irregularity, and motion characteristics for fire smoke detection [2]. Their study proposed a robust AdaBoost (RAB) classifier to improve training and classification accuracy. Prema and Suresh suggested a smoke detection method by analyzing its texture features [3]. They employed the gray level co-occurrence matrix (GLCM) and gray level run length matrix (GLRL) to extract the spatial structures of a local binary pattern code map. Wang et al. have developed a fast detection method for early fire smoke [4]. Their proposed algorithm depends on the smoke's color and diffusion characteristics, and it counts the number of pixels in each candidate smoke region. Their results indicated that a better detection speed and accuracy had been achieved for the proposed algorithm, and the proposed algorithm can deal with the problem of moving elements, such as pedestrians and cars. Gagliardi and Saponara established a video-based smoke detection technique, advanced video smoke detection (AdViSED), for early warning in antifire surveillance systems [5]. Their method can quickly detect fire smoke and send an alarm signal using a Kalman estimator, color analysis, image segmentation, blob labeling, geometrical features analysis, and M of N decisor.

Traditional DIP methods for smoke detection face several challenges. They are often sensitive to changes in lighting and confuse other atmospheric phenomena, such as fog or dust, with smoke, resulting in high false alarms. They might struggle in dynamic or cluttered environments, and they might not effectively distinguish between different types of smoke in terms of color, density, texture, or movement. Some of these algorithms require regular calibration for optimal performance and depend on fixed thresholds, which might be less adaptive. As a result, these traditional DIP methods are now being gradually outpaced by more advanced techniques such as DL. In addition to eliminating the need for domain experts to perform feature engineering, DL can directly use the original image as the input source in the numerical model, which promotes the analysis and application to be diversified and popularized.

The DL has been developed dramatically with three primary applications: smoke classification, smoke detection, and smoke segmentation. Firstly, smoke classification uses models such as convolutional neural network (CNN) to determine whether smoke is present in an image or video. It broadly categorizes input into smoke or non-smoke. Shees et al. presented a light-weight neural network, FireNet-v2, which was deployed to the Raspberry Pi 3B for smoke and flame detection [6]. The CNN structure has 14 layers, mainly composed of convolution layers, average pooling layers, and dropout layers. The results showed that the accuracy rates were 98.43% and 94.95% in Foggia's dataset and the FireNet dataset, respectively. Cao et al. introduced a spatio-temporal cross network (STCNet) involving a spatial pathway to extract texture features and a temporal pathway to capture smoke motion information to recognize industrial smoke emissions [7]. The two-stream network architecture includes the spatial path to extract the characteristics of smoke appearance, background information, and the temporal path to extract the features of the moving target. Identification was performed with the bidirectional feature fusion of these two paths in the form of element addition on the feature maps of different scales. Wu et al. proposed a two-stage real-time video smoke detection method based on dense optical flow and CNN [8]. The features were extracted by inputting images filtered through the suspicious smoke areas (SSA) module and using the optical flow of the blue channel (OFBC) for motion detection. Then, the ResNet-34 model was utilized as the main architecture and video frame weight optimization method for smoke detection. Almeida et al. established a lightweight CNN-based wild fire-smoke detection program— EdgeFireSmoke [9]. EdgeFireSmoke was installed in edge computing devices that analyzed aerial photographs from unmanned aerial vehicles (UAVs) and closed-circuit television (CCTV) surveillance systems. Secondly, smoke detection identifies the location and size of a smoke area within a scene using bounding boxes. Zhang et al. used Faster R-CNN to identify fire smoke in the forest, such that the complex manual feature extraction process can be avoided for traditional video smoke detection methods [10]. Their study utilized synthetic smoke images with the insertion of real smoke or simulated smoke into the forest background in Faster R-CNN, due to the lack of training data. Saponara et al. presented a fire and smoke detection model using YOLOv2, based on a real-time video in antifire surveillance systems [11]. YOLOv2 is designed with light-weight neural network architecture to account for the requirements of embedded platforms. Huo et al. proposed a deep separable CNN for smoke detection [12]. The convolution path, space pyramid pool (SPP) module, and depthwise-separable convolution were added on YOLOv4 in order to extract image features more effectively, to enhance features of small targets, and to reduce network parameters, respectively. The results showed that their method was more sensitive to early smoke, and the accuracy rate was 98.5%. Chen et al. associated the mixture Gaussian model with YOLOv5 to detect smoke for textile workshops [13]. Mixture Gaussian model was used to extract suspected smoke areas, and the target recognition ability of YOLOv5 was improved by adding with an adaptive attention module. Wang et al. detected smoke using YOLOv5m [14]. The mosaic enhancement method, the dynamic anchor box module, and the attention mechanism were added on YOLOv5m in order to solve the problems of small numbers of smoke training samples, prior inaccurate anchor box information, and unbalanced feature maps in different scales, respectively. The mean average precision (mAP) of their model was 4.4% higher than those of the traditional DL algorithm. Finally, smoke segmentation delves deeper than classification and detection, marking every pixel as either part of the smoke or not. It offers a detailed mask to outline the boundary of smoke. Yuan et al. attempted to monitor the occurrence of smoke in various scenarios [15]. They generated a dataset of synthesized smoke using virtual smoke synthesis techniques and, then, performed smoke segmentation using a two-stream fully convolutional network (FCN). Their method is capable of segmenting smoke, to some extent, in different scenes. However, it still produces some false alarms when dealing with clouds and fog. Gupta et al. introduced a pixel-level algorithm, considering both spatial and temporal features, for wildfire video detection in forests [16]. They used a dynamic optimal frame (DOF) module,

combined with ResNet-50, on the input video data to reduce the amount of haze in video frames and extract smoke feature frames. Then, Mask R-CNN was employed to perform a region of interest (ROI) operation to achieve automatic labeling.

Previous studies on DL training tasks usually necessitate large, labeled datasets and assume that the training and testing data have the same feature space and distribution, making their implementations both time-consuming and low-generalization. Due to domain shift, smoke classifiers, detectors, or segmentors trained for one context struggle to generalize to other unfamiliar smoke scenarios, leading to a significant decline in prediction performance [17]. This is a contrast to domain adaptation (DA) algorithms, which are designed to adjust seamlessly to new scenarios. DA aims to extract domain knowledge by using original domain data and transferring the knowledge learned to new domains to improve the problem of insufficient labeled data and reduce computing costs. Pan and Yang divided the transfer learning context into three categories, according to the data labeling status: inductive transfer learning, transductive transfer learning, and unsupervised transfer learning [18]. The DA algorithm is one of the extended applications of transfer learning. Cross-domain feature learning defines the training data and test data as the source domain and the target domain, respectively. It is mainly used in the source domain data that have been labeled, as well as the target domain. The objective of DA is to solve the domain shift problem between the source domain and the target domain. A deep DA network, based on statistical differences for video smoke detection, has been presented by Xu et al. [19]. The network structure used AlexNet as the backbone for feature extraction and CORAL as the loss function, combining the gradient reversal layer (GRL) for domain confusion, to calculate the difference in feature distribution between domains. Their experimental results show that the highest detection rate of the algorithm in identifying real smoke was 94.70%. Zhou et al. proposed an unsupervised DA smoke detection algorithm based on so-called multilevel feature cooperative alignment and fusion (MCAF). Cooperative domain alignment was performed on a multilevel feature extraction network in order to reduce domain difference. Multilevel feature fusion modules were embedded at different depths of the network to enhance the representation ability for small objects [17]. Comparing with YOLOv3, YOLOv5s, and Faster-RCNN on cross-domain tasks among RFdataset, SFdataset, and True_smoke, the MCAF is more suitable for smoke detection. Xu et al. proposed a DA smoke detector based on single-shot multi-box detector (SSD) and multi-scale deep convolutional neural network (MSCNN) [20]. An adversarial training algorithm was designed to adapt the detector trained on synthetic smoke images to real scenes of the USTC_SmokeRS dataset. The highest mean location accuracy (mLA) was only 64.17%. Mao et al. proposed a feature-level domain adaptation method that incorporated adversarial discriminative domain adaptation (ADDA) with DeepCORAL [21]. This method was adopted to reduce the domain shift between the synthetic and real smoke images. The analyzed dataset came from State Key Laboratory of Fire Science, USTC and from Yuan Feiniu, where an accuracy of 97.39% was obtained.

Multiple production lines are usually set up for an injection molding factory to achieve continuous, parallel manufacturing and multi-functional production. The application of visual fire alarms combined with CNN in industrial smoke detection is often limited by the production line's available type, mechanical structure, and site setting. It would be time-consuming to retrain a new CNN model whenever a new production line is created. The present study aims to develop a system that combines DIP, CNN, and DA for smoke detection through a fixed video camera, observing injection molding production lines. The training data are obtained by synthesizing the smoke images with the field background of the injection molding production. The DIP is employed to extract features for the synthesized images. The frame difference is the establishment of the attention mechanism of the smoke detection algorithm, which is aiming to realize the continuous tracking of the feature pixels of moving objects by CNN. Due to the similarity of the injection molding production lines in the smart factory, the DA method is utilized for different

injection molding production lines without retraining the CNN algorithm. The research's positioning is listed in below:

1. Real smoke creation, extraction, and synthesis: The injection molding production line usually maintains a stable production status, so it is difficult to obtain real smoke images in practice. Repeated experiments have shown that artificial smoke, produced by burning "smoke cakes", has the characteristics of low cost, low risk, and high similarity to real smoke. By producing artificial smoke against a green screen as a background, the DIP method is used to extract the smoke image and combine it with the background of the injection molding area. This creates an image of the smoke in the injection molding area, which is then used as training data for the smoke classification CNN algorithm.
2. Smoke classification via CNN algorithm: We embedded a smoke classification CNN algorithm in visual fire alarms above the production line. By leveraging frame differences with DIP and motion detection, as well as using an attention mechanism, the CNN tracks moving object features, pinpointing smoke characteristics.
3. Integration of CNN with DA: The injection molding production lines in the smart factories discussed in this research have similar features. Using the DA method means that the CNN algorithm does not need to be retrained for different injection molding lines. Applying the CNN algorithm to an injection molding line only requires the synthesis of smoke images with the scene of the injection molding production line to complete the auto-annotation and algorithm training steps. This saves the effort and time of re-labeling smoke images and retraining the algorithm.

2. Research Methods

In this chapter, we introduce the videos of scene and smoke acquisition, as well as the generation processes. A real time smoke detection based on deep domain adaptation has been developed for different injection molding machines at the same plant site.

2.1. Videos of Scene and Smoke Acquisition

There were two fixed webcams that recorded the videos of the present study. The devices in the videos are two injection molding machines, the conveyor belt, and the surrounding environment, as shown in Figure 1a,b. The webcams were located at a height of 2.5 m in front of the machines. Virtual smoke videos were generated to simulate abnormal situations at the plant site. The smoke videos of the source domain are mainly obtained from the smoke videos generated by burning the smoke cake with a green background, as presented in Figure 1c. On the other hand, the smoke in the target domain is simulated with on-site smoke placed on the injection molding machines, as shown in Figure 1d, as the actual state of the abnormal condition.
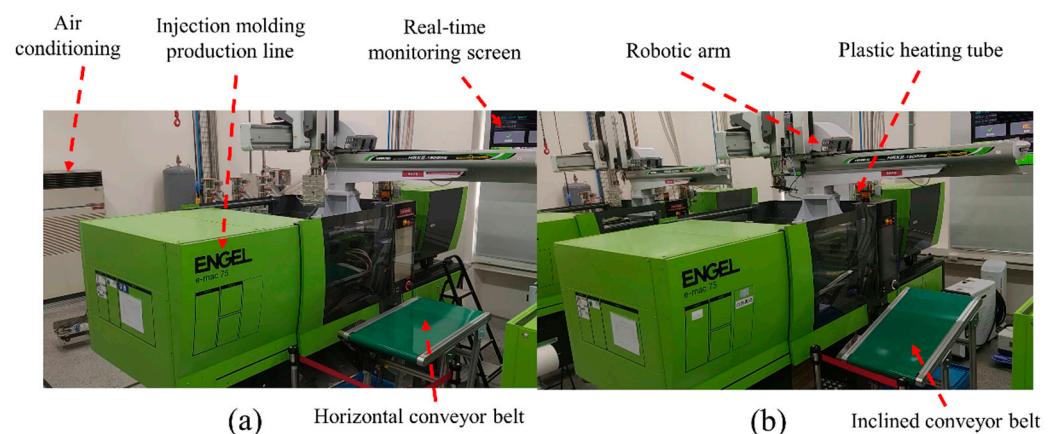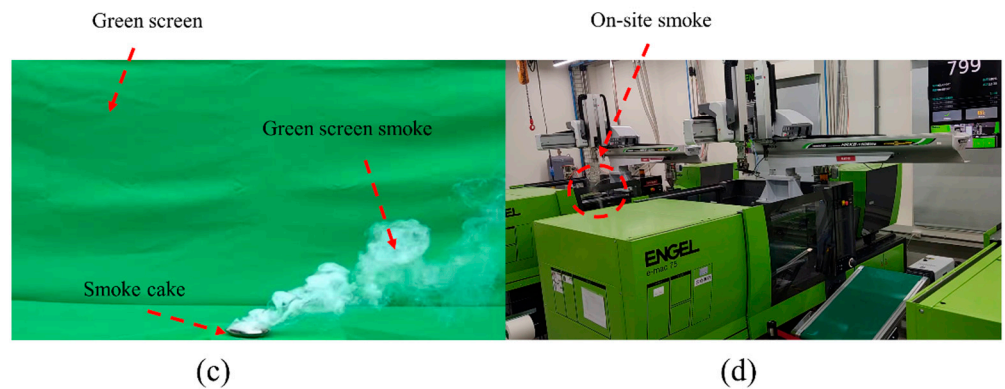

Figure 1. *Cont.*

(c)   (d)

Figure 1. Monitoring scenes and video generation: (**a**) the injection molding machine A; (**b**) the injection molding machine B; (**c**) green screen smoke; (**d**) on-site smoke.

2.2. Video Pre-Processing

The procedures of the video pre-processing are depicted in Figure 2. In this study, the smoke videos with green screen were converted from RGB to HSV. To remove the green screen background and to preserve green screen smoke, the range of Hue of was set between 138 and 140, the Saturation was 0, and the Value was 300.
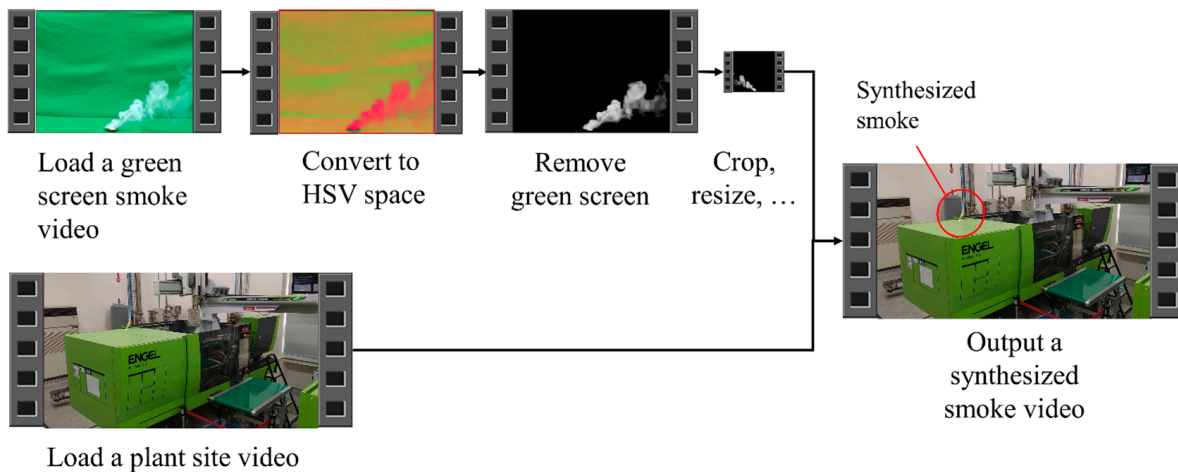
**Figure 2.** Procedure of synthesized smoke video generation.

The synthesized smoke videos for injection molding machines were generated through image augmentation to perform ROI scaling, offset, cropping, grayscale processing, and HSV contrast conversion for the background-removed smoke videos. The purpose is to enrich the training data set and prevent the occurrence of training overfitting. However, because the smoke is translucent, the synthesized videos were unnatural if we used the element-wise addition operator directly on the two videos. Here, it is necessary to consider the brightness value of each pixel of the smoke image and measure the ratio of the background image to be synthesized. The image synthesis formula for each channel is as shown in Equation (1).

$$\text{Synthesize}_{i,j} = (255 - \text{Smoke}_{i,j})/255 \times \text{PlantSite}_{i,j} + \text{Smoke}_{i,j} \qquad (1)$$

where i and j represent the x and y coordinates, respectively, Synthesize is an RGB image that represents the resulting synthesized image, PlantSite is also an RGB image that represents the background of the plant site, and Smoke is a Value map of the green screen smoke image from the HSV color space.

2.3. Smoke Detection with CNN

The present investigation employs CNN image classification to identify whether smoke appears in the scene. The structure of the CNN is shown in Figure 3. The structure is mainly composed of two-step filters: a convolutional layer with the filter size of 3×3 and a ReLU activation function with the window of a 2×2 maximum pooling layer. It should be noted that the video was converted into a series of $1 \times 255 \times 3$ tensors, instead of the originally captured images, as the input of CNN.

(a)   (b)   (c)   (d)
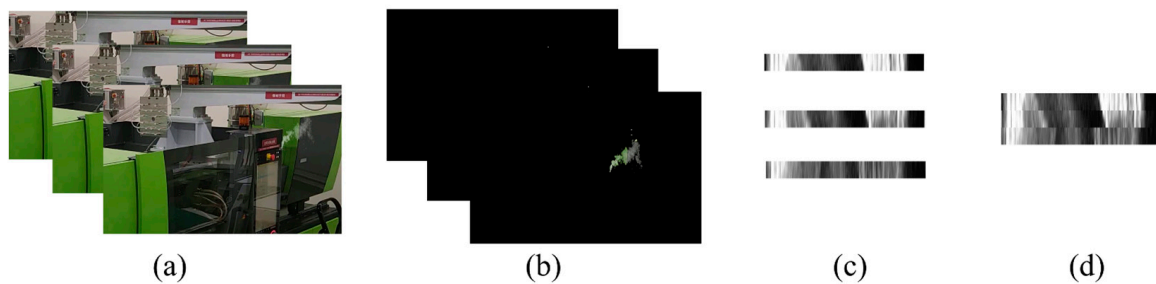
Figure 3. Images obtained before CNN identification: (**a**) original image; (**b**) residual mass; (**c**) gray histogram image; (**d**) image from feature fusion.

There are three methods utilized before CNN image classification. The Gaussian mixture models (MOG) method generates a residual mask, as shown in Figure 3b, to capture the motion pixels. The MOG method is a mechanism for spatial attention that can guide CNN's eyes to focus on the moving object instead of the background. Later, a $1 \times 255 \times 1$ gray histogram image for the pixel values was created, as presented in Figure 3c, where 1×255 represents the frequency of occurrence of gray levels ranging from 1 to 255, and the 1 of the third dimension is the number of frames. Feature fusion was employed to extract the moving pixels and convert them into three consecutive grayscale values. As shown in Figure 3d, through this multi-frame feature fusion process, the dimension of the CNN input image becomes $1 \times 255 \times 3$, which could further capture dynamic features.

2.4. Domain Adaptation

In the present investigation, the source domain is one machine, and the target domain is another. Ganin et al. proposed the domain adversarial training of neural networks (DANN) structure, as shown in Figure 4, for domain adaptation [22]. As presented in Figure 4, DANN mainly comprises three networks: feature extractor (f), label predictor (y), and domain classifier (D). After images are input from either the source domain or the target domain, the feature extractor network extracts features using two successive convolution layers and max-pooling layers. The output feature maps are then flattened into 1D feature vectors. The 1D feature vectors generated by feature extractor are finally given to the label predictor and domain classifier for two classification tasks simultaneously. Both the label predictor and domain classifier, especially, are classifiers whose network structures are two successive dense layers for the purpose of binary classification. During the training process, the labeled data from source domain and the unlabeled data from target domain are mixed.

The loss functions, corresponding to the inverse transfer, are expressed as below:

$$\theta_{\mathcal{F}} = \theta_{\mathcal{F}} - \alpha \left(\frac{\partial L_y^i}{\partial \theta_{\mathcal{F}}} - \lambda \frac{\partial L_D^i}{\partial \theta_{\mathcal{F}}} \right) \tag{2}$$

$$\theta_y = \theta_y - \alpha \frac{\partial L_y^i}{\partial \theta_y^i} \tag{3}$$

$$\theta_D = \theta_D - \alpha \frac{\partial L_D^i}{\partial \theta_D^i} \tag{4}$$

where $\theta_{\mathcal{F}}$, θ_y, and θ_D represent feature, label, and domain mapping parameters, L_y^i and L_D^i represent the loss function of i-th sample label prediction y and domain classification D, respectively, a is the learning rate, and -1 is the GRL coefficient whose purpose is to control the degree of domain confusion.
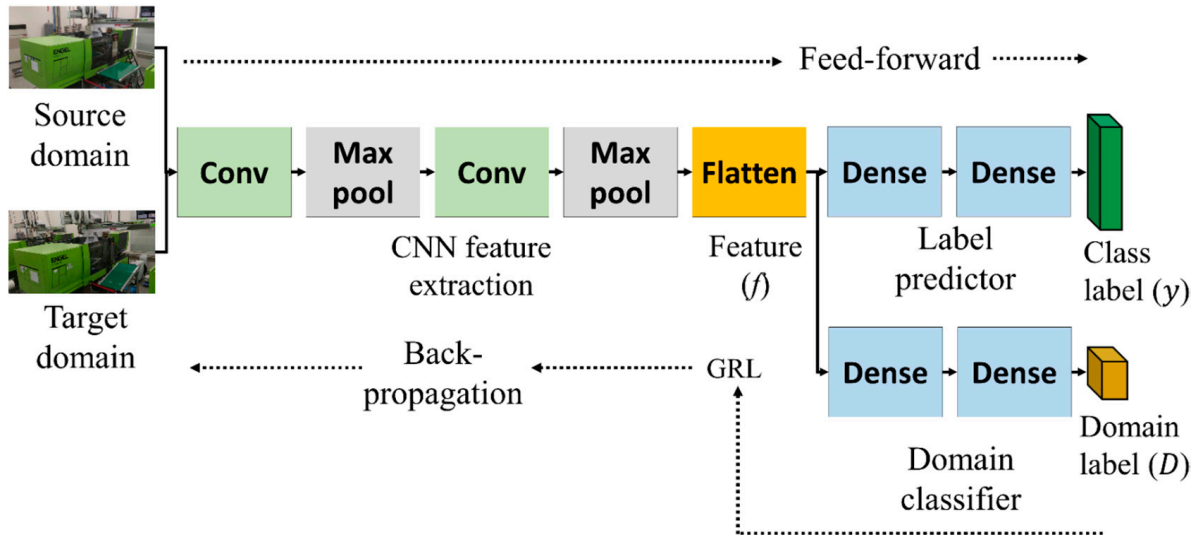

Figure 4. Structure of DANN.

When the backward transfer of updating the weights was completed for DANN, the loss function of DANN could be obtained through Equation (5):

$$E(\hat{\theta}_{\mathcal{F}}, \hat{\theta}_y, \hat{\theta}_D) = \sum_{\substack{i=1\ldots n \\ D_i=0}} L_y^i(\hat{\theta}_{\mathcal{F}}, \hat{\theta}_y) - \lambda \sum_{i=1\ldots n} L_D^i(\hat{\theta}_{\mathcal{F}}, \hat{\theta}_D) \tag{5}$$

where $\hat{\theta}_y$ and $\hat{\theta}_D$ are the loss functions of the label predictor and domain classifier, and they will converge through Equations (6) and (7), respectively.

$$\hat{\theta}_y = \min_{\theta_y} L_y \tag{6}$$

$$\hat{\theta}_D = \max_{\theta_D} L_D \tag{7}$$

The feature extractor $\hat{\theta}_{\mathcal{F}}$ can be calculated from Equation (8).

$$\hat{\theta}_{\mathcal{F}} = \min_{\theta_{\mathcal{F}}} L_y - L_D \tag{8}$$

As the image is input, the label predictor can be correctly classified, and the domain classifier cannot distinguish the source of the image at the same time.

2.5. Model Evaluation Metrics

The confusion matrix is used for the evaluation of the performance of the model in the prediction of classification problems. In addition to calculate accuracy rate, sensitivity, specificity, and F1-score are also employed as the main evaluation metrics to explore the DA algorithm in the smoke detection task because the smoke image is inhomogeneous compared with the whole background.

3. Experiment Results

In this chapter, the video data sets have been processed, and the MOG method was employed to detect moving objects in the frames. Domain adaptations with three structures—deep domain confusion (DDC) [23], domain separation network (DSN) [24],

and DANN [22]—were evaluated. Various hyper-parameter experiments have been carried out, and optimization of the model is also performed in the present study.

3.1. Preprocessing and Features of Video Datasets

In the investigation, fixed webcams were used for video recording. The source of the smoke video is divided into two parts: green screen smoke and on-site smoke. The former is taken outdoors with burning smoke cakes with a green screen, while the latter is released around the injection molding machines. These videos are preprocessed through motion detection to generate a time-series image data set. The size of the image is 1920 × 1080 pixels. The scenarios include the smoke along with the molding machines, light source, people movement, and environmental interferences. The presence of smoke in the video is considered an abnormal situation.

There are 2196 images prepared for machine A. In the images for machine A, the training samples included 730 frames of normal condition, 445 frames of synthesized smoke, and 372 frames of on-site smoke, while the test samples had 299 frames of normal condition and 350 frames of on-site smoke. In addition, there are 2238 images prepared for machine B. In the images for machine B, the training samples include 730 frames of normal condition, 370 frames of synthesized smoke, 440 frames of on-site smoke, while test samples include 370 frames of normal condition and 328 frames of on-site smoke.

Figure 5 presents the original image with corresponding moving pixel positions extracted by MOG, residual masks, a grayscale histogram, and a line chart for various scenarios. It should be noted that the injection molding machine shown in Figure 5(a1) is in a static scenario. The moving scenarios include dynamic background, light source, human walking, and smoke situations corresponding to Figure 5(a2–a5), respectively. The dynamic background situation is caused by the opening and closing of the molding machine when in production; the light source situation comes from the sunlight outside of the plant and the light reflection by humans passing in front of the machine; the human walking situation derives from the moving objects of the people on-site walking around the machine; the smoke situation is the main target of the present study.

Additionally, Figure 5(b1–b5) are the frame difference images obtained by MOG, which can extract the position of the moving pixels, and Figure 5(c1–c5) are the corresponding residual masks, which can be extracted from the foreground, corresponding to the moving pixels, achieving the effect of the attention mechanism. Figure 5(d1–e5) are the quantified grayscale histogram and line chart of the residual mask, respectively. The quantified grayscale histogram and line chart of the residual mask indicates that the characteristics of the static condition are very different from those of moving conditions. In the moving conditions, the characteristics of the smoke condition are very different from the other three conditions as well.
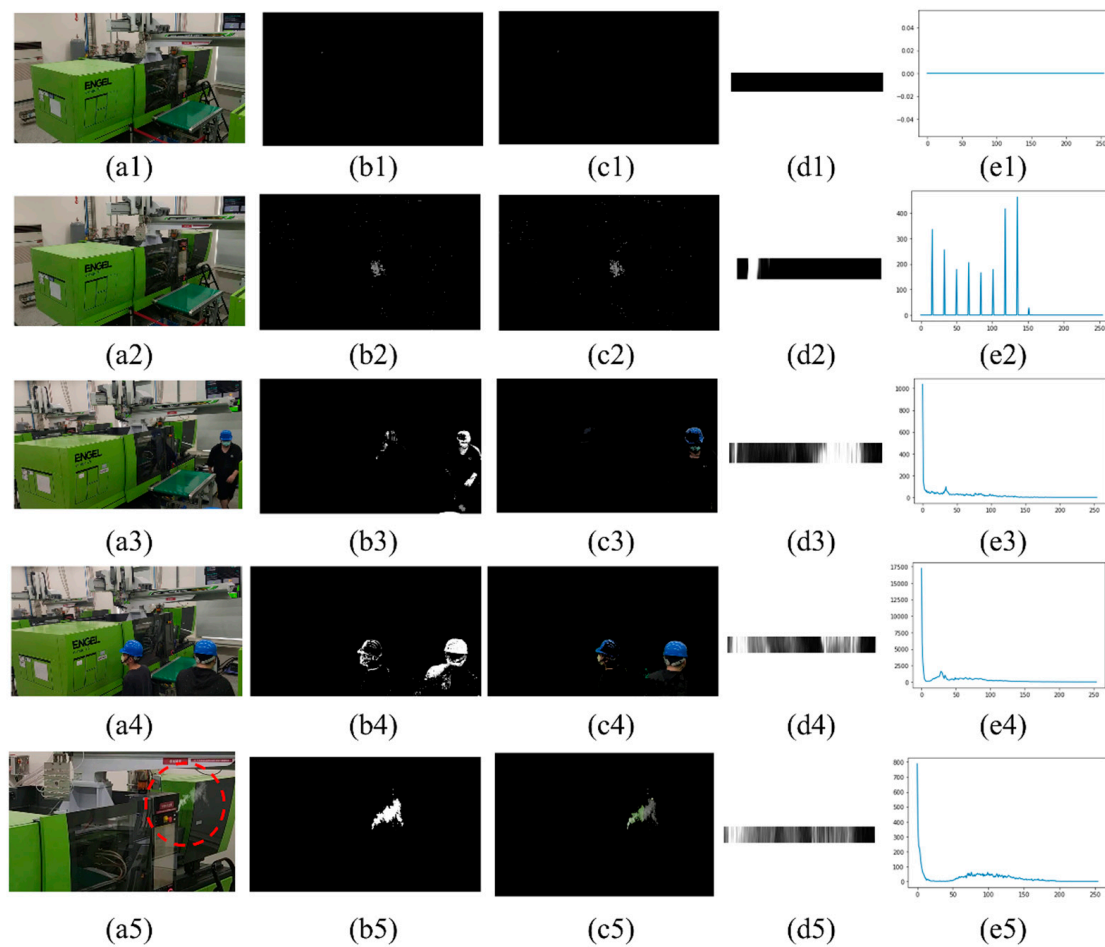
Figure 5. A display of residual masks and the corresponding grayscale histogram under various scenarios with MOG extraction. (**a1**–**a5**), respectively, describe the normal, dynamic background, light source, human walking, and smoke situation examples; (**b1**–**b5**) are the corresponding moving pixel positions extracted by MOG; (**c1**–**c5**) are the corresponding residual masks; (**d1**–**d5**) are the corresponding images after the residual mask is extracted by the grayscale histogram; (**e1**–**e5**) correspond to the line chart.

3.2. Model Spot Checking

In this section, the domain adaptation with three different domain migration modes—DDC, DSN, and DANN—were discussed. The DDC structure was developed based on the concept of difference, while the DSN structure was developed based on the concept of reconstruction, and the DANN structure was developed based on the concept of confrontation. These domain migration structures were employed in this model spot checking of smoke detection in two injection molding machines on the site, aiming to obtain an effective domain adaptation effect by finding a suitable domain adaptation algorithm for the data of the present study.

As demonstrated in Figure 6, scenario (A→A) means training with the data of machine A and, then, testing with the data of machine A. The classification accuracy rate with CNN [25] is 70.20%, while scenario (B→A) is using data of machine B for training and testing with the data of machine A; however, the classification accuracy rate with CNN is only 39.20%. Therefore, a better classification accuracy rate with DDC, DSN, and DANN is obtained as 63.02%, 64.23%, and 65.80% respectively. In addition, scenario (B→B) indicates training with the data of machine B and, then, testing with the data of machine B. The classification accuracy rate with CNN is 69.60%, while scenario (A→B) is using data of machine A for training and testing with the data of machine B; however, the classification

accuracy rate with CNN is only 47.80%. Again, better classification accuracy rates with DDC, DSN, and DANN are obtained as 63.32%, 65.74%, and 68.20%, respectively.
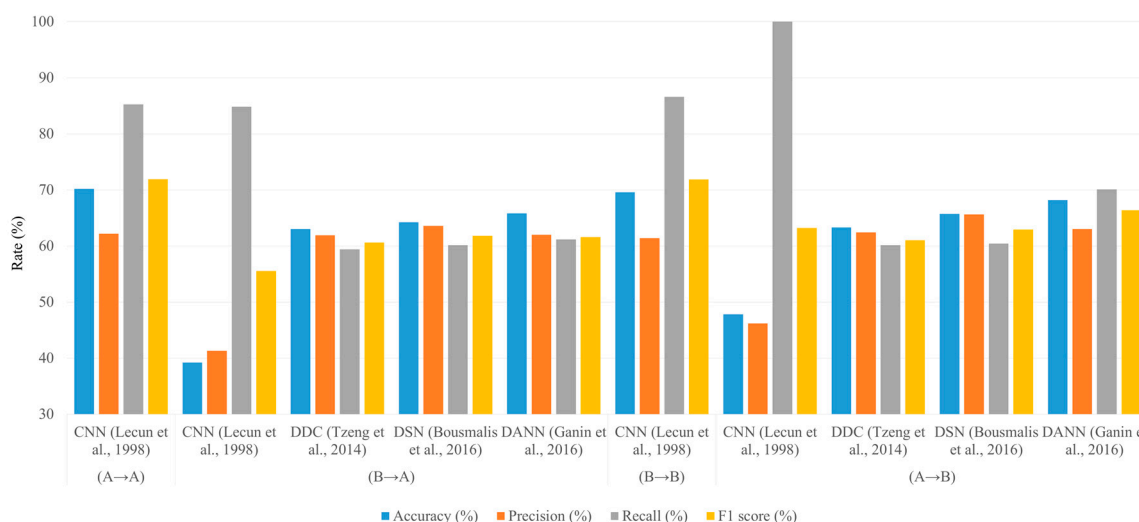
Figure 6. Result of model spot checking [22–25].

Through the experiment mentioned above, two brief conclusions are reached. The first is the migration and generalization of the domain adaptation model between unlabeled target domain data and source domain data, thereby improving the prediction of the model on both source domain data and target domain data. The second is that the employment of DANN has a relatively better classification accuracy rate. Therefore, DANN is chosen as the basis algorithm of smoke detection for domain adaptation.

3.3. Ablation Study

The DANN model is based on the mechanism of confrontation to achieve the effect of domain confusion. It uses the mixed features extracted from CNN to increase the accuracy of the label predictor to identify the classes of the input data and make the domain classifier unable to distinguish whether the frame is from the source or target domain, leading to the inference of the model for the data in the unlabeled target domain. The number of iterations required for DANN model training has a great influence on the mixture of domain features. The number of iterations required by the DANN model during training has a great impact on the mixing of domain features. CNN requires a certain number of iterations for the confusion of feature distribution in feature learning in the initial stage, so the balance between the feature extractor, label predictor, and domain classifier must be considered as the number of iterations increases. In this section, the early stopping mechanism and reconfirm mechanism of the smoke detection algorithm are discussed. The performance of the model has been improved through the optimization mechanism and fine-tuning methods. The results of the smoke detection algorithm in this study are summarized as follows:

3.3.1. Design of Experiment (DOE)

The DOE is utilized in the present investigation to optimize the hyper-parameters of the smoke detection domain adaptation algorithm mentioned previously. As presented in Table 1, the proposed smoke detection domain adaptation algorithm is a five-factor, two-level L32(25) full factorial experiment. The description of each factor and level are listed as the following:

- Motion detection: MOG is used to extract a moving object in the image, and its level is divided into default mode and custom parameter mode. The default mode of the MOG's parameters are: 500 frames of background image, threshold of variation as 16, and dynamically adjusted learning rate; the custom mode of the MOG parameters

reduces the number of background images to 100 due to interference of the noise of the light source, and it increases the threshold of variation to 25. In addition, it is necessary to have the MOG stop updating the background modeling as MOG detects enough moving pixels, i.e., the learning rate is set to 0 with consideration of the lower diffusion of smoke. Then, MOG will return to the dynamic adjustment until the abnormal state is resolved.

- Image sources: The residual mask obtained from the original video through moving object detection is an input image source for CNN. However, since the number of moving object pixels in an image is not so obvious compared to the background, the residual mask was further converted into a grayscale histogram, ignoring background in the grayscale histogram, such that CNN can learn the information of moving pixels more effectively. In this way, CNN can be more capable of expressing feature information as another input image source for CNN.

- Information fusion: The data set in this research is time-series data, and the multi-frame information fusion technology of the input image can make CNN evolve from only considering spatial information to considering spatial-temporal information. In the experiment, single-frame and three-frame are set as the level configuration of this factor.

- Confusion coefficient: DANN uses GRL technology to confuse the data distribution of the source and target domain when the error is back-propagated. In the initial stage of training, it is necessary to use iteration for the domain classifier such that the domain classifier is not able to determine the data from the domain sources. Then, the domain adaptation is achieved. In the experiment, the confusion coefficient is 0.001 and 0.01 as the level configuration, respectively.

- Optimizer: The choice of optimizer is closely related to the updating of model weights. The default optimizer of DANN for weight learning is SGD. In the experiment, Adam is also employed as an alternative optimizer for weight learning so that the optimizer would further consider gradient speed adjustment, learning rate adjustment, and parameter deviation correction in the analysis. This optimizer makes the updating process of the model weight more stable.

Table 1. Hyper-parameter and level setting of DOE.

Factors	Motion Detection	Image Sources	Information Fusion	Confusion Coefficient	Optimizer
Level 1	Default	Residual mask	Single-frame	0.001	SGD
Level 2	Custom	Grayscale histogram	Multi-frame	0.01	Adam

There are 32 full factorial experiments carried out for scenario (A→B) and scenario (B→A) training. The classification accuracy of these experimental data are recorded individually. The A→B represents that the real time video of machine A was used as the source domain, and then, it was adapted to the target domain video of machine B and vice versa. Later, the experimental data are plotted as the main effect plot, as shown in Figure 7, to display the suggested hyper-parameter combination between factors and levels. It is concluded that the best combination for the proposed model is custom mode for the motional detection, grayscale histogram for the image source, three-frame for the information fusion, 0.001 for the confusion coefficient, and the Adam algorithm for the optimizer.
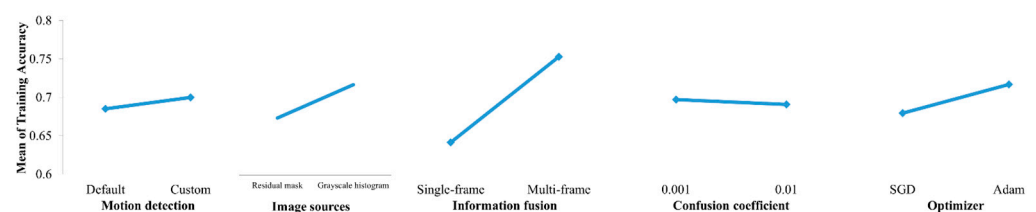

Figure 7. Main effect plot for various hyper-parameters.

3.3.2. Early Stopping Mechanism

Early stopping is a regularization method. If the classification accuracy rate does not improve for ten consecutive times, the training process will stop. Generally, the early stopping mechanism can avoid over-fitting on the training data set. In the experiment, in the analysis, the early stopping mechanism is used to improve the training robustness of DANN. The experimental results show that, even though the scenarios of machine A and machine B are similar, they are still regarded as two independent domains, and the number of model iterations of each scenario needs to be fine-tuned to reduce the gradient effectively and stabilize the accuracy rate during training.

Figure 8(a1,b1) are the classification accuracy curves of the DANN model under different epochs in the scenario (A→B) and scenario (B→A), while Figure 8(a2–b3) are the corresponding loss function curves. It is found that, when the classification accuracy rate curve is stable, the values of loss function drop sharply; however, sudden increases of the function curve have been observed after the stabilization of the classification accuracy rate curve, leading to the instability of network training. Therefore, an early stop mechanism is provided to avoid this problem.
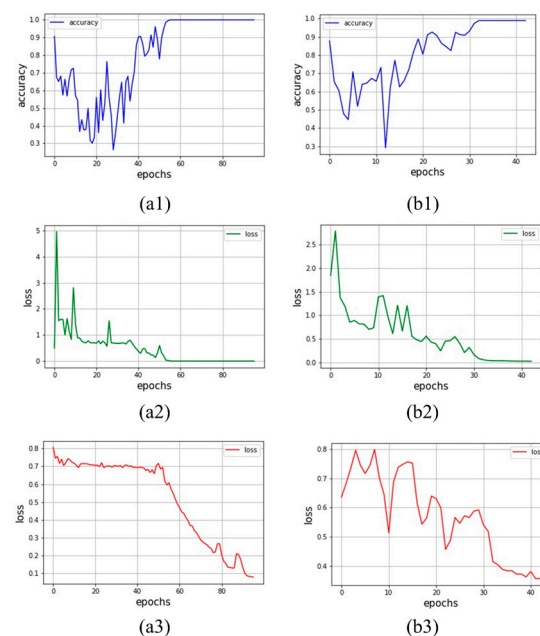
Figure 8. Line chart of the evaluation measurements. (**a1**,**b1**) are classification accuracy rates of DANN during training in scenario (A→B) and scenario (B→A); (**a2**,**b2**) are the loss functions of the label predictor during training in scenario (A→B) and scenario (B→A); (**a3**,**b3**) are the loss functions of the domain classifier during training in scenario (A→B) and scenario (B→A).

3.3.3. Reconfirm Mechanism

Using the prediction information with single frame as the result, the model prediction would be too sensitive, resulting in high false alarm rate and missed alarm rate. Therefore, the result is reached five consecutive times for smoke.

3.3.4. Results of the Ablation Experiment

Results of the ablation experiment are presented in Figure 9. For the scenario (B→A), the naïve DANN [22] achieved accuracy of 66.54%, precision of 65.87%, recall of 58.05%, and F1 score of 61.71%. By adding the DOE optimization mechanism, there was a noticeable improvement across all metrics: 86.73% accuracy, 91.44% precision, 80.76% recall, and 85.77% F1 score. The significant improvements in various metrics are mainly due to the DOE's ability to suggest a better set of hyper-parameters. In addition to recommending a smaller confusion coefficient for GRL and selecting Adam as the weight update algorithm,

MOG is advised by DOE to establish a reference background based on short-term information. When smoke is determined to be present, the update of the reference background is immediately halted. This allows the reference to adjust, in real-time, with the current status of the background, and it prevents the inclusion of smoke data in the reference. Furthermore, the DOE suggests using a grayscale histogram instead of a residual mask as input, which more effectively quantifies the features of moving pixels. Moving pixels make up a small proportion of an image, so the residual mask is a sparse 2D image. CNNs usually handle global information better and find it harder to focus on local information. Therefore, it is necessary to help CNNs perform a certain degree of feature transformation. Finally, the DOE recommends using multi-frame information as input, allowing the classifiers to learn the dynamic information of moving pixels. Incorporating both the DOE optimization mechanism and the early stopping mechanism resulted in accuracy of 86.82%, precision of 86.83%, recall of 89.08%, and F1 score of 87.94%. The early stopping mechanism does not significantly improve overall accuracy, but it is more helpful for balancing precision and recall. Finally, combining the DOE optimization mechanism, early stopping mechanism, and the reconfirm mechanism, the performance peaked with accuracy of 93.17%, precision of 92.2%, recall of 98.56%, and F1 score of 95.27%. The reconfirm mechanism can be seen as a form of ensemble prediction, making the model more conservative and reliable when outputting classification predictions. Each prediction requires continuous and multiple confirmations before a conclusion is reached. From Figure 9, it is evident that the scenario (A→B) also follows the same trend. Compared with DANN without introducing any mechanism, the accuracy, precision, recall, and F1 score have been improved from 67.01%, 66.81%, 65.11%, and 65.95% to 91.35%, 86.37%, 96.94%, and 91.35%, respectively, with the addition of all mechanisms.
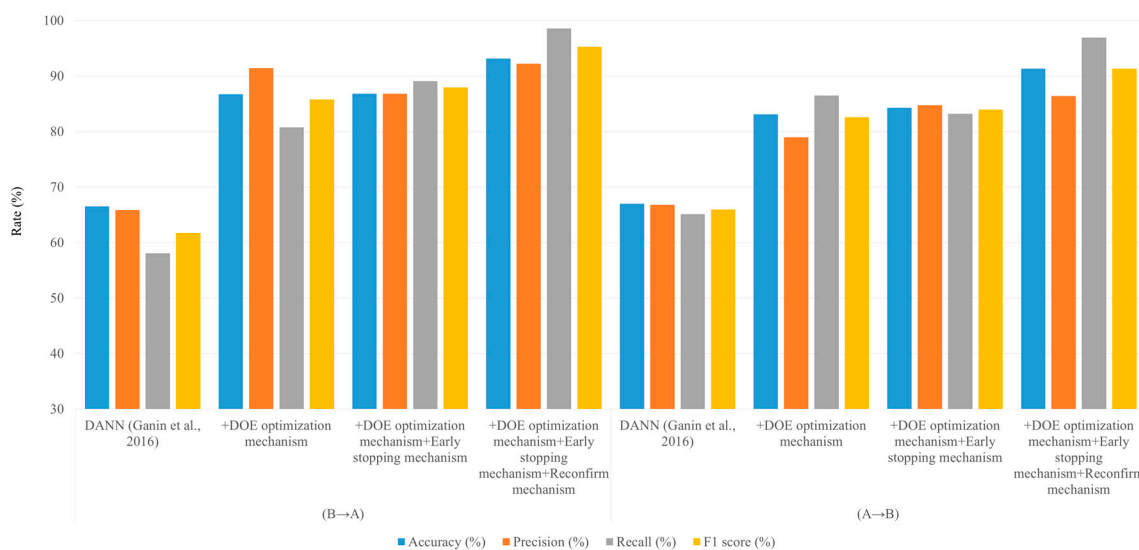

Figure 9. Ablation study in testing proposed DANN model [22].

3.4. Feature Distributions of DANN

The main objective of domain adaptation is to achieve domain confusion. t-distributed stochastic neighbor embedding (t-SNE) is a non-linear dimensionality reduction algorithm that can project the high-dimensional data into low-dimensional graphics for visualization. The feature distributions can be observed through t-SNE visualization. Figure 10 displayed domain and class feature distributions for scenario (A→B) and scenario (B→A). These feature distributions of the DANN are used to explore the variations of the domain feature and class feature distribution after DANN learning.
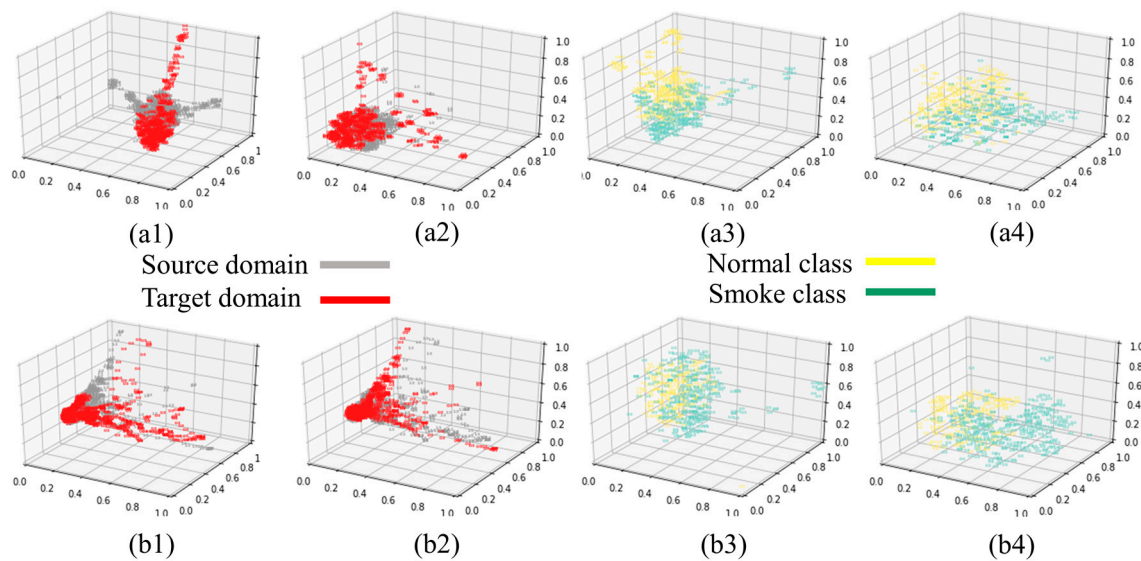
Figure 10. Feature distribution with t-SNE visualization: (**a1**–**a4**) are the domain and class before training for scenario (A→B), and (**b1**–**b4**) are domain and class after training for scenario (B→A).

From the perspective of domain features, the domain feature distributions of machine A and machine B, before training, are demonstrated in Figure 10(a1) and (b1), respectively, while the domain feature distributions after training are presented in Figure 10(a2,b2). It can be observed that, before training, there is a noticeable difference in the appearance of feature distributions between the source and target domains for both scenarios, and there is a certain distance between them. It is found that, after training, the domain features between these two domains have been confused, and the transfer learning of the model is realized from the domain invariant features extraction.

From the perspective of class features, Figure 10(a3) and (b3), respectively, represent the class feature distributions, before training, in two scenarios. It is observed that the feature distributions for normal and smoke classes are mixed together and have not yet been clearly distinguished. After learning by DANN, the class feature distributions are depicted in Figure 10(a4,b4). Apparently, the feature distributions for normal and smoke classes show dramatic differences. After enhancing the distinction between the feature distributions of each class, one can correctly differentiate between the normal and smoke states.

3.5. Demonstration of Model Inference

The optimized smoke detection domain adaptation model presented in the previous section is employed to infer the test data of injection molding machine A and B, respectively. The information of a moving object would appear in the upper left corner of the picture. For example, "Normal" in the green color, in the upper left corner of the picture, means that the model determines that the condition of the scenario is normal. If "Smoke" in the red color appears in the upper left corner of the picture, it means that smoke has been detected by the model, and the alarm should be set off. There are five scenarios depicted in Figure 11(a1–b5). Figure 11(a1,b1) represent the machines in standby mode, with no detected motion pixels. Therefore, those images are naturally considered as "Normal". Figure 11(a2–b4) represent the machines in operation mode, light reflection, and human walking, where motion pixels are detected at the corresponding region and are correctly considered as "Normal". Figure 11(a2–b4) represent smoke around machines where motion pixels are detected at the smoke region and are correctly considered as "Smoke". The proposed method is able to determine the "Normal" for the four normal scenarios correctly, regardless of the machine motion and surrounding interference. In addition, the smoke detection domain adaptation model can be used to identify smoke with migrating data from the source domain to the target domain.
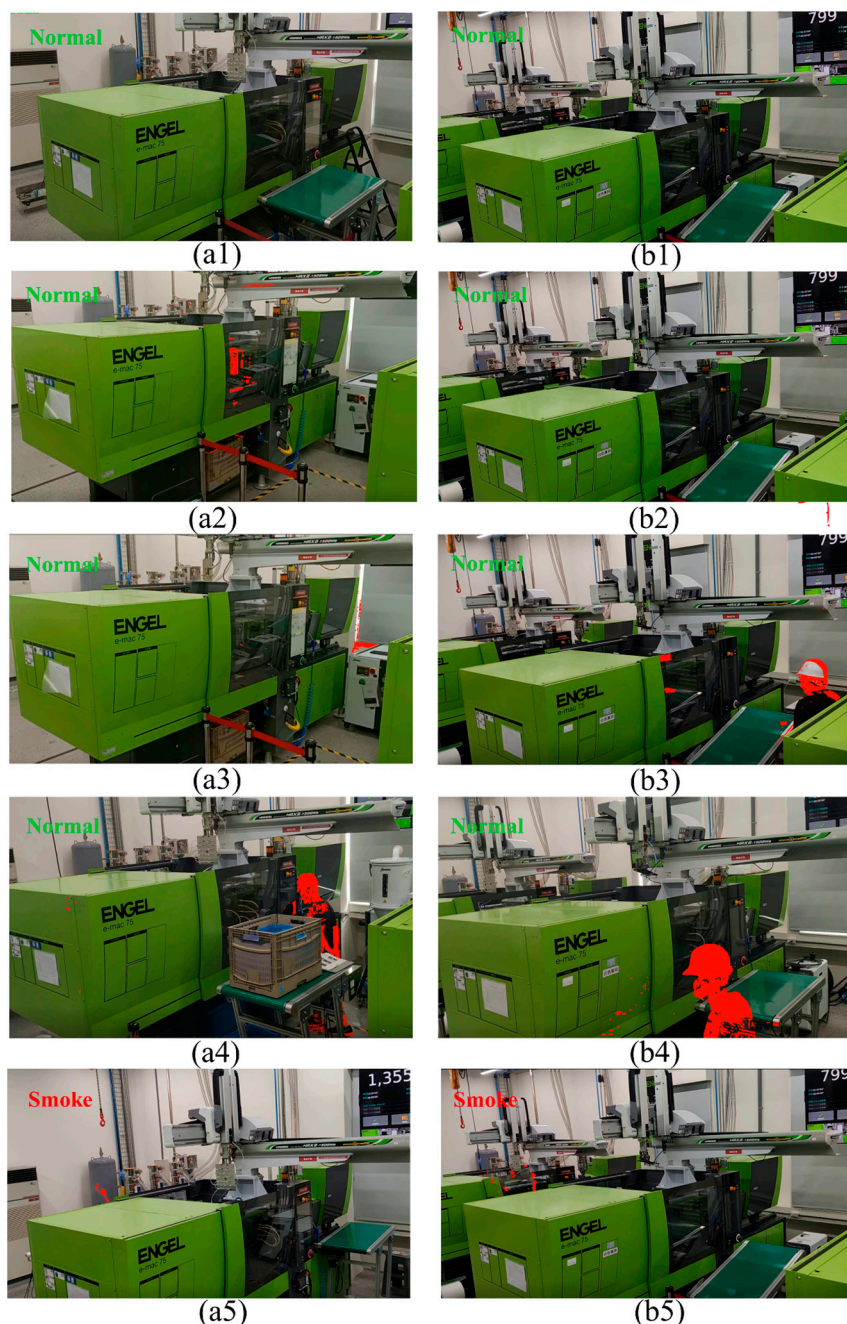
Figure 11. Results of inference with the proposed smoke detection algorithm. (**a1,b1**) are machines in standby mode; (**a2,b2**) are machines in operation; (**a3,b3**) are light reflections of machines due to light sources; (**a4,b4**) are humans walking around machines; (**a5,b5**) are smoke around machines.

4. Conclusions

In recent years, intelligent smart fire protection has become an indispensable trend. DL has been widely used in visual fire alarms. Due to insufficient training data and labeling issues, it takes a lot of time and effort for supervised learning, as well as manpower and costs. The deep domain adaptation network is a scheme of transfer learning, and it learns the labeled source domain data and unlabeled target domain data at the same time. This network can understand the two sets domain data simultaneously through domain confusion. Inference can be made to achieve the automatic labeling of data in the target field.

In the present analysis, an adaptive algorithm for smoke detection, based on motion detection combined with DANN, has been proposed. This algorithm is applied to two

injection molding machines on the production line, and these machines have a certain degree of similarity to each other. The images of the dynamic background of two machines, light sources, and people walking on-site are collected. The smoke images are obtained by burning smoke cakes outdoors with a green background. After a series of background removal, augmentation, and image synthesis, these processed smoke images are input in the proposed algorithm through the training process of domain transfer. Brief summaries are listed below:

1. Domain adaptation of smoke: The work of two injection molding machines adapting to each other is realized by using DANN network. There is no need of labeling for target domain data, so the model possesses the characteristics of automatic labeling, two domain knowledge, and the function of no retraining for the target domain. The model can be used as pre-trained weights for transfer learning, in other injection molding domains, to save manpower and time.
2. Hyper-parameter DOE optimization experiment: The best combination of hyper-parameters is found using custom MOG parameter, three-frame grayscale histogram fusion, a confusion coefficient setting of 0.001, and optimizer setting of Adam. The average classification accuracy rate of the two scenarios can reach 84.90%, compared to that of 65% for the original size image as input.
3. The effects of the early stopping mechanism and the reconfirm mechanism: The training process and prediction of DANN is stabilized with the combination of these two mechanisms. More importantly, these two mechanisms help to increase the classification accuracy rate to more than 90%.

This study has two limitations. First, all the smoke is produced from smoke cakes, not from the malfunction of injection molding machines. Second, the method is currently only applied to indoor settings, and its domain adaptation effectiveness for outdoor scenarios remains uncertain. Currently, DA used in this study is limited to smoke classification application. Future research can expand DA to applications in smoke detection and smoke segmentation, further highlighting the benefit of saving annotation costs in the target domain.

Author Contributions: Conceptualization, C.-Y.W. and C.-M.C.; methodology, S.-H.C. and M.-J.Y.; software, C.-H.K. and Y.-T.C.; validation, C.-Y.W. and C.-M.C.; formal analysis, Y.-T.C. and C.-H.K.; data curation, J.-S.L. and J.-K.L.; writing—original draft preparation, J.-H.J. and S.-H.C.; writing—review and editing, J.-H.J. and S.-H.C.; supervision, K.F.-R.L. and M.-J.Y.; project administration, K.F.-R.L.; funding acquisition, K.F.-R.L. and S.-H.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

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

References

1. Mutar, A.F.; Dway, D.H.G. Smoke detection based on image processing by using grey and transparency features. *J. Theor. Appl. Inf. Technol.* **2018**, *96*, 6995–7006.
2. Wu, X.; Lu, X.; Leung, H. A Video Based Fire Smoke Detection Using Robust AdaBoost. *Sensors* **2018**, *18*, 3780. [CrossRef] [PubMed]
3. Prema, C.E.; Suresh, S. Local binary pattern based hybrid texture descriptors for the classification of smoke images. *Int. J. Eng. Res. Technol.* **2019**, *7*, 1–7.
4. Wang, H.; Zhang, Y.; Fan, X. Rapid Early Fire Smoke Detection System Using Slope Fitting in Video Image Histogram. *Fire Technol.* **2020**, *56*, 695–714. [CrossRef]
5. Gagliardi, A.; Saponara, S. AdViSED: Advanced Video SmokE Detection for Real-Time Measurements in Antifire Indoor and Outdoor Systems. *Energies* **2020**, *13*, 2098. [CrossRef]
6. Shees, A.; Ansari, M.S.; Varshney, A.; Asghar, M.N.; Kanwal, N. FireNet-v2: Improved lightweight fire detection model for real-time IoT applications. *Procedia Comput. Sci.* **2023**, *218*, 2233–2242. [CrossRef]

7.   Cao, Y.; Tang, Q.; Lu, X. STCNet: Spatiotemporal cross network for industrial smoke detection. *Multimed. Tools Appl.* **2022**, *81*, 10261–10277. [CrossRef]
8.   Wu, Y.; Chen, M.; Wo, Y.; Han, G. Video smoke detection base on dense optical flow and convolutional neural network. *Multimed. Tools Appl.* **2020**, *80*, 35887–35901. [CrossRef]
9.   Almeida, J.S.; Huang, C.; Nogueira, F.G.; Bhatia, S.; de Albuquerque, V.H.C. EdgeFireSmoke: A Novel Lightweight CNN Model for Real-Time Video Fire–Smoke Detection. *IEEE Trans. Ind. Inform.* **2022**, *18*, 7889–7898. [CrossRef]
10. Zhang, Q.-X.; Lin, G.-H.; Zhang, Y.-M.; Xu, G.; Wang, J.-J. Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images. *Procedia Eng.* **2018**, *211*, 441–444. [CrossRef]
11. Saponara, S.; Elhanashi, A.; Gagliardi, A. Real-time video fire/smoke detection based on CNN in antifire surveillance systems. *J. Real-Time Image Process.* **2020**, *18*, 889–900. [CrossRef]
12. Huo, Y.; Zhang, Q.; Jia, Y.; Liu, D.; Guan, J.; Lin, G.; Zhang, Y. A Deep Separable Convolutional Neural Network for Multiscale Image-Based Smoke Detection. *Fire Technol.* **2022**, *58*, 1445–1468. [CrossRef]
13. Chen, X.; Xue, Y.; Zhu, Y.; Ma, R. A novel smoke detection algorithm based on improved mixed Gaussian and YOLOv5 for textile workshop environments. *IET Image Process.* **2023**, *17*, 1991–2004. [CrossRef]
14. Wang, Z.; Wu, L.; Li, T.; Shi, P. A Smoke Detection Model Based on Improved YOLOv5. *Mathematics* **2022**, *10*, 1190. [CrossRef]
15. Yuan, F.; Zhang, L.; Xia, X.; Wan, B.; Huang, Q.; Li, X. Deep smoke segmentation. *Neurocomputing* **2019**, *357*, 248–260. [CrossRef]
16. Gupta, T.; Liu, H.; Bhanu, B. Early wildfire smoke detection in videos. In Proceedings of the 2020 25th International Conference on Pattern Recognition, Milan, Italy, 10–15 January 2021; pp. 8523–8530.
17. Zhou, F.; Wen, G.; Ma, Y.; Wang, Y.; Ma, Y.; Wang, G.; Pan, H.; Wang, K. Multilevel feature cooperative alignment and fusion for unsupervised domain adaptation smoke detection. *Front. Phys.* **2023**, *11*, 1136021. [CrossRef]
18. Pan, S.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
19. Xu, G.; Zhang, Y.; Zhang, Q.; Lin, G.; Wang, J. Deep domain adaptation based video smoke detection using synthetic smoke images. *Fire Saf. J.* **2017**, *93*, 53–59. [CrossRef]
20. Xu, G.; Zhang, Q.; Liu, D.; Lin, G.; Wang, J.; Zhang, Y. Adversarial Adaptation From Synthesis to Reality in Fast Detector for Smoke Detection. *IEEE Access* **2019**, *7*, 29471–29483. [CrossRef]
21. Mao, J.; Zheng, C.; Yin, J.; Tian, Y.; Cui, W. Wildfire Smoke Classification Based on Synthetic Images and Pixel- and Feature-Level Domain Adaptation. *Sensors* **2021**, *21*, 7785. [CrossRef]
22. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
23. Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; Darrell, T. Deep domain confusion: Maximizing for domain invariance. *arXiv* **2014**, arXiv:1412.3474.
24. Bousmalis, K.; Trigeorgis, G.; Silberman, N.; Krishnan, D.; Erhan, D. Domain separation networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 343–351.
25. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]