big data and cognitive computing

Article

A Novel Method for Improving Baggage Classification Using a Hyper Model of Fusion of DenseNet-161 and EfficientNet-B5

Mohammed Ali Saleh [1], Mohamed Abdouh [2] and Mohamed K. Ramadan [2,*]

[1] Department of Computers and Systems Engineering, Faculty of Engineering, Helwan University, Cairo 11795, Egypt; mohamed.saleh@hq.helwan.edu.eg

[2] Department of Computer Science, Faculty of Computer Science, Nahda University in Beni Suef, Beni Suef 62521, Egypt; mohamed.abdo@nub.edu.eg

* Correspondence: khaled.abotyra@nub.edu.eg

Abstract: In response to rising concerns over crime rates, there has been an increasing demand for automated video surveillance systems that are capable of detecting human activities involving carried objects. This paper proposes a hyper-model ensemble to classify humans carrying baggage based on the type of bags they are carrying. The Fastai framework is leveraged for its computational prowess, user-friendly workflow, and effective data-cleansing capabilities. The PETA dataset is utilized and automatically re-annotated into five classes based on the baggage type, including Carrying Backpack, Carrying Luggage Case, Carrying Messenger Bag, Carrying Nothing, and Carrying Other. The classification task employs two pretrained models, DenseNet-161 and EfficientNet-B5, with a hyper-model ensemble that combines them to enhance accuracy. A "fit-one-cycle" strategy was implemented to reduce the training time and improve accuracy. The proposed hyper-model ensemble has been experimentally validated and compared to existing methods, demonstrating an accuracy of 98.6% that exceeds current approaches in terms of accuracy, macro-F1, and micro-F1. DenseNet-161 and EfficientNet-B5 have achieved accuracy rates of 95.5% and 97.3%, respectively. These findings contribute to expanding research on automated video surveillance systems, and the proposed model holds promise for further development and use in diverse applications.

Keywords: deep learning; hyper model; pre-processing pipeline; baggage type; Fastai framework

Citation: Saleh, M.A.; Abdouh, M.; Ramadan, M.K. A Novel Method for Improving Baggage Classification Using a Hyper Model of Fusion of DenseNet-161 and EfficientNet-B5. *Big Data Cogn. Comput.* **2024**, *8*, 135. https://doi.org/10.3390/bdcc8100135

Academic Editors: Min Chen, Robail Yasrab and Md Mostafa Kamal Sarker

Received: 21 August 2023
Revised: 10 November 2023
Accepted: 14 December 2023
Published: 11 October 2024

1. Introduction

The rapid growth of urbanization and increasing global concerns over public safety and security have led to the need for enhanced video surveillance systems [1,2]. These systems play a crucial role in monitoring public spaces, securing critical infrastructures, and detecting criminal or suspicious activities [3]. A key aspect of these surveillance systems is the capability to accurately detect and classify human activities, particularly those involving carried objects such as baggage [4,5]. This is essential in various situations, including airport security, public transportation, and crowded events, where the identification of individuals carrying potentially hazardous or prohibited items is of the utmost importance [6].

Deep learning techniques, specifically convolutional neural networks (CNNs), have demonstrated significant success in various computer vision tasks, including object detection, classification, and human activity recognition [7,8]. However, there remains a need for further improvement in the detection and classification of human-carried baggage to ensure higher accuracy and adaptability to real-world scenarios [9,10]. The detection and classification of human-carried baggage using deep-learning techniques face challenges such as occlusion, varied appearances, and real-time processing requirements [11]. Accurately detecting and classifying baggage is further complicated by factors like background clutter, motion blur, and limited annotated data [12]. Addressing these challenges is crucial for enhancing video surveillance systems and ensuring public safety [13,14]. In response to these pressing needs, we present an innovative approach that leverages advanced deep-learning techniques. The significance of our work lies in its potential to dramatically enhance the detection and classification capabilities of video surveillance systems in real-world scenarios, particularly those involving human-carried baggage. Our research promises substantial contributions to public safety and security across various contexts, ranging from airport security to public transportation and crowded events. Beyond the immediate application in surveillance, our innovations resonate profoundly within the broader landscape of computer vision. Our pioneering methodologies, from applying hyper model application to baggage-based human classification to innovative preprocessing strategies, aim to drive the evolution of computer vision methodologies and their myriad applications. To this end, we focused on the following contributions:

1. We applied hyper models for the first time to the problem of classifying humans carrying baggage based on bag types.
2. We developed a reliable hyper model that can classify humans, with or without baggage, into five classes based on the baggage type.
3. We introduced a pre-processing pipeline that includes increasing the image contrast, applying a sharpening filter, adjusting image brightness and saturation, and removing noise to improve model performance.
4. We automatically re-annotated the PETA dataset with direct information about the baggage type using a custom Python script.
5. We implemented the innovative ‘fit-one-cycle’ policy method to reduce the number of epochs and iterations required for our model to handle large-scale data.

This paper is organized as follows: Section 2 presents a review of related work on deep learning techniques for detecting and classifying human-carried baggage; Section 3 describes the methodology, including the dataset, preprocessing steps, and the proposed hyper model ensemble; Section 4 presents the experimental results and discussion; and finally, Section 5 presents the conclusion with a discussion on potential future work.

2. Related Work

Recent years have seen the increasing application of deep learning techniques in various computer vision tasks, including object detection, classification, and human activity recognition [4]. This section provides an overview of the related work on deep learning techniques for detecting and classifying human-carried baggage and critically examines the gaps that persist.

2.1. Human Activity Recognition

Human activity recognition (HAR) is a critical component in video surveillance systems and has been widely studied over the past few years [15]. Several deep learning-based HAR techniques have been proposed, such as using convolutional neural networks (CNNs) [16], recurrent neural networks (RNNs) [17], and long short-term memory (LSTM) networks [18]. However, despite their efficacy in recognizing basic human actions like walking and running, these techniques often struggle when it comes to intricate actions involving carried objects. The granularity of such actions, particularly when objects are partially obscured, makes them challenging to detect and classify accurately.

2.2. Object Detection and Classification

Deep learning-based object detection and classification techniques have gained popularity due to their ability to identify objects in images and videos with high accuracy [19]. Techniques like Faster R-CNN [20], the Single Shot Multi-Box Detector (SSD) [21], and You Only Look Once (YOLO) [22,23] have been developed to address these challenges. While these techniques have shown promise in detecting various objects, their performance in classifying human-carried baggage is still limited due to challenges such as occlusion, the diverse appearances of baggage based on angles and lighting, and the need for instantaneous processing that limits their efficacy [11].

Ahammed et al. [24] leveraged deep learning and YOLOv3 for the real-time detection of unattended baggage and owner identification, achieving notable accuracy. This work underscores the essential role of precise detection and the classification of human-carried baggage in enhancing security, an area to which our prior work has made substantial contributions. We discuss these in the following subsection.

2.3. Human-Carried Baggage Detection and Classification

Numerous studies have focused on detecting and classifying human-carried baggage using deep-learning techniques. Chen et al. [25] proposed a two-stage framework for baggage recognition, where a CNN is initially employed to detect humans in images, followed by a second CNN to classify the baggage type. However, this approach relies on separate models for human detection and baggage classification, which may not be efficient for real-time applications.

To address the challenges in human-carried baggage detection and classification, researchers have explored various techniques, such as attention mechanisms [26], multi-scale feature learning [27], and temporal information integration [28]. Nonetheless, these methods continue to face limitations in terms of their accuracy, adaptability to real-world scenarios, and processing time.

Wahyono et al. [29] developed a framework using custom CNN layers and transfer learning based on the human viewing direction for classifying objects carried by humans. Despite its efficacy, this approach faces limitations due to occlusions, similarities in human appearance, and challenges in identifying front-facing backpack carriers, requiring further refinement.

In our previous work [30], we addressed this protruding problem by proposing a model for classifying humans carrying baggage. Our approach involved utilizing the pretrained DenseNet-161 architecture and employing a “fit-one-cycle policy” strategy to minimize the training time and enhance accuracy. The model demonstrated high precision and outperformed existing techniques, achieving an accuracy range of 96–98.75% for binary classification and 96.67–98.33% for multi-class classification. Notably, our method proved to be effective in detecting baggage when it was heavily obscured or indistinguishable from other objects. Moreover, our study focused on the direction in which humans were looking, not the type of bag they carried.

In summary, while prior studies have showcased significant advancements in human-carried baggage detection and classification, they also highlight a clear gap: the need for a method that ensures heightened accuracy, real-time processing, and precise bag type classification. This paper presents a novel approach, leveraging advanced deep-learning techniques, particularly hyper-models, to address these challenges and augment the efficiency of video surveillance systems.

3. Methodology

This section provides a detailed description of the methodology used in this study, including dataset preparation, data preprocessing, and the development of the proposed hyper model.

3.1. Dataset Description and Preparation

In this sub-section, a thorough explanation is provided regarding the dataset used in this study, including the dataset re-annotation process and how data augmentation techniques were applied to enhance the dataset.

3.1.1. Dataset Description

The dataset utilized in this study is the PETA dataset [31], comprising 19,000 annotated images of pedestrians with various attributes, such as clothing, gender, age, and object interactions. This dataset was selected based on its extensive coverage, including diverse settings and rich annotations, making it well-suited for detecting and classifying human-carried baggage, which is a critical task in security applications.

3.1.2. Dataset Re-Annotation

To align the PETA dataset with our research objectives, we conducted a comprehensive dataset re-annotation process by categorizing the images into five distinct classes based on the type of baggage being carried. These five classes included Carrying Backpack, Carrying Luggage Case, Carrying Messenger Bag, Carrying Nothing, and Carrying Other.

To automate this process, we developed a custom Python script that used the attributes of the dataset to map the original annotations to the new baggage-based classes. This approach allowed us to efficiently adapt the PETA dataset to our research needs, resulting in a more focused dataset that was better suited for the task of baggage classification.

Table 1 illustrates the distribution of images across the five newly annotated classes, which provides valuable insights into the composition of our refined dataset. It displays the number of images for each class and the total number of images after re-annotation. Additionally, Figure 1 displays a representative sample of the re-annotated dataset and displays a sample of images from each of the five classes, providing visual examples of the types of baggage that fall into each class.

Table 1. Distribution of images across the five classes after re-annotation.

Class	Number of Images
Carrying Backpack	3385
Carrying Luggage Case	389
Carrying Messenger Bag	4666
Carrying Nothing	4801
Carrying Other	5759
Total	19,000

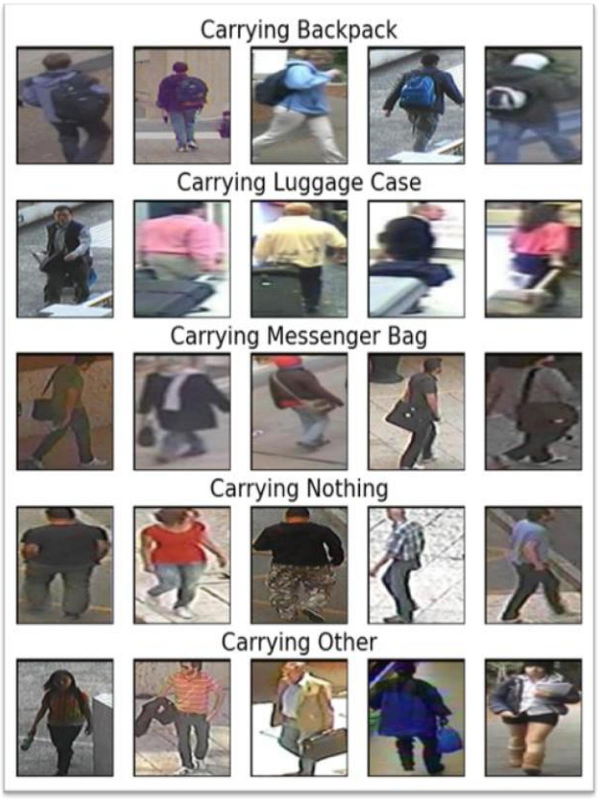


Figure 1. A representative sample of the re-annotated dataset showcasing the five distinct classes.

3.1.3. Data Augmentation

In this study, data augmentation techniques [32] were applied only to the "Carrying Luggage Case" class, involving the generation of additional training samples to increase the dataset size from 389 to 2000 images. Various techniques, such as image rotation, flipping, zooming, and translation, were used to create a more diverse dataset for this class. This process aimed to enhance the model's ability to accurately recognize and classify "Carrying Luggage Case", which is a critical task in security applications. While the other categories did not require data augmentation, the overall dataset was still augmented to ensure consistency in the dataset's size and diversity across all classes. By selectively applying data augmentation techniques to the "Carrying Luggage Case" class, we were able to significantly improve the model's performance in detecting and classifying this class while maintaining consistency in the dataset.

3.2. Data Preprocessing

This subsection covers the Image Resizing and Image Pre-processing Pipeline, which are essential components of pipelines for image classification.

3.2.1. Image Resizing

To ensure consistency and compatibility with deep-learning models, all images in the dataset were resized to a uniform dimension of 224 × 224 pixels. This size was chosen as it is a common input size for deep learning architectures, such as DenseNet and EfficientNet.

3.2.2. Image Pre-Processing Pipeline

To optimize the model's performance, a sophisticated pre-processing pipeline was implemented for the images [33]. This pipeline encompasses several crucial steps aimed at refining image quality and enhancing the overall capabilities of the model. The key stages involved in this pipeline are as follows:

- Contrast enhancement: This step focuses on increasing the contrast of the image, which amplifies the distinction between various objects and their background. Enhanced contrast facilitates improved object recognition and localization.
- Sharpening Filter Application: By applying a sharpening filter, the edges and boundaries of objects within the image were emphasized. This process aids in better feature extraction and helps the model identify and differentiate between objects more effectively.
- Brightness and saturation adjustment: Adjusting image brightness and saturation is vital for enhancing color information and overall visibility. This step allows for more accurate color-based object recognition, thereby contributing to the model's performance.
- Noise reduction: The process of removing noise through Gaussian blur and non-local means of denoising is crucial to minimize the impact of small artifacts and image noise. By lowering noise levels, this model can concentrate on relevant features, leading to a more precise analysis.

Figure 2 provides examples of images that have undergone the aforementioned preprocessing steps, demonstrating the improvements in image quality and the potential benefits for the performance of our proposed model.

3.3. Development of Proposed Model

This sub-section describes the development of the proposed model for baggage detection and classification, which involves the selection and fusion of the DenseNet-161 and EfficientNet B5 architectures, the design of a hyper model, and the application of the fit-one-cycle training method. Figure 3 illustrates the architecture of the hyper model. Additionally, we outline the evaluation criteria used to assess the model's performance.

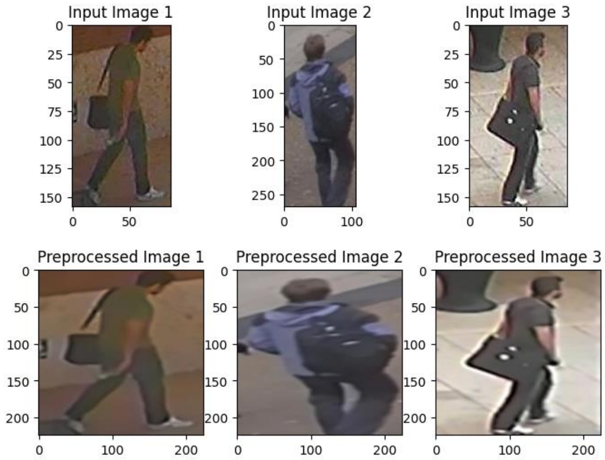


Figure 2. Examples of pre-processed images that show improved quality and potential benefits for model performance.

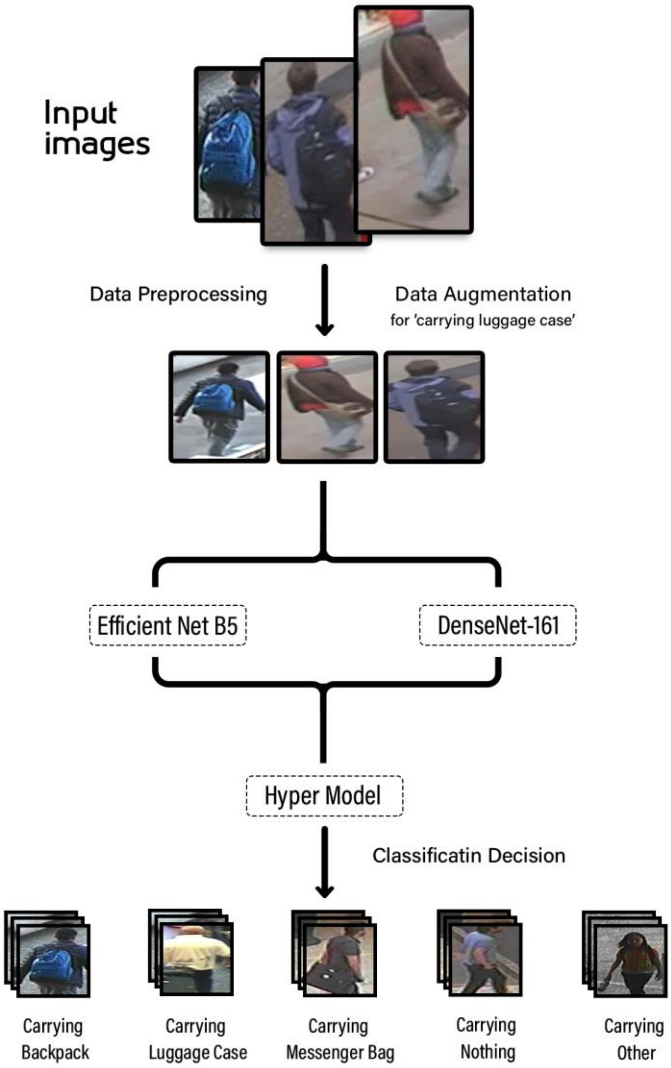


Figure 3. The processing steps in our suggested model's approach.

3.3.1. Densenet-161 Architecture

DenseNet-161 [34] is a variant of the DenseNet architecture, which is well-known for its excellent performance in image classification tasks. DenseNet-161 consists of 161 layers and is characterized by dense connections between layers, wherein each layer receives input from all the preceding layers. This design promotes efficient gradient flow and feature reuse, leading to improved performance while using fewer parameters compared to traditional CNN architectures.

This architecture is organized into multiple dense blocks, each containing several convolutional layers. These dense blocks, which are integral components of the DenseNet-161 architecture, are interconnected through transition layers. These layers play a pivotal role in managing feature map dimensions, effectively reducing computational complexity and thereby enhancing the efficiency of the overall system. Figure 4 demonstrates the connection mechanism among various convolutional layers within a DenseNet block. These feed-forward connections increase the total layer count from L to L × (L + 1)/2. In this study, the DenseNet-161 architecture was employed as one of the base models for the proposed hyper model.

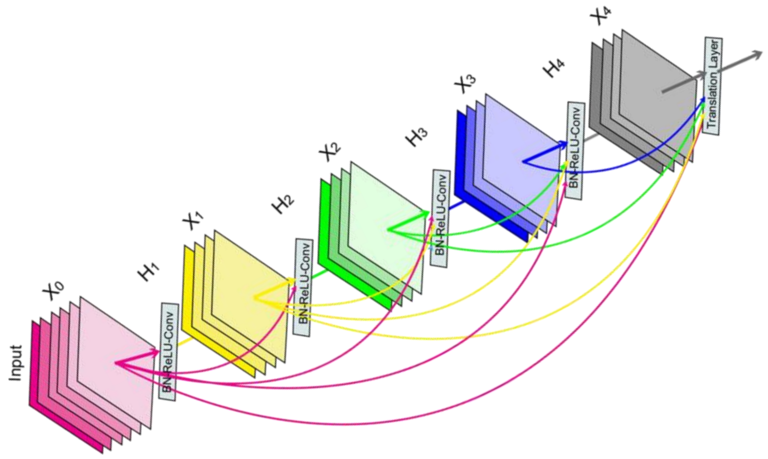


Figure 4. A dense block where each layer uses all the feature maps from the previous layers as input. (The pink arrow represents the initial input or starting point of the process. The colored layers transitioning from pink to grey signify various stages of processing. The arrows labeled "BN → ReLU → Conv" denote a sequence of operations common in neural networks or deep learning, where "BN" stands for Batch Normalization, "ReLU" is the Rectified Linear Unit activation function, and "Conv" refers to a Convolutional layer. These are integral components of a Convolutional Neural Network (CNN). The 'Translation Layer' is indicated as the concluding stage in the diagram. The arrows depict the direction of data flow and the sequence of operations within the process).

3.3.2. EfficientNet B5 Architecture

EfficientNet [35] is a family of CNN architectures that exhibit exceptional performance in various computer vision tasks, including image classification. The EfficientNet B5 is a specific instance of the EfficientNet architecture that is designed to strike a balance between model efficiency and accuracy. It employs a unique compound scaling method that simultaneously scales the depth, width, and resolution of the network to achieve optimal performance.

EfficientNet B5 embodies a highly advanced model that strategically extends the bedrock established by EfficientNet B0 and serves as its fundamental architectural backbone. The B5 version is an upscaled variant of the base model with increased depth, width, and input resolution. This scaling enables the network to capture more complex features and improve its overall performance. Moving beyond the design and workings of EfficientNet B5, it is crucial to highlight its application within the scope of our research. Figure 5 illustrates the architecture of EfficientNet B5. In our study, the EfficientNet B5 architecture was also employed as one of the base models for the proposed hyper model.

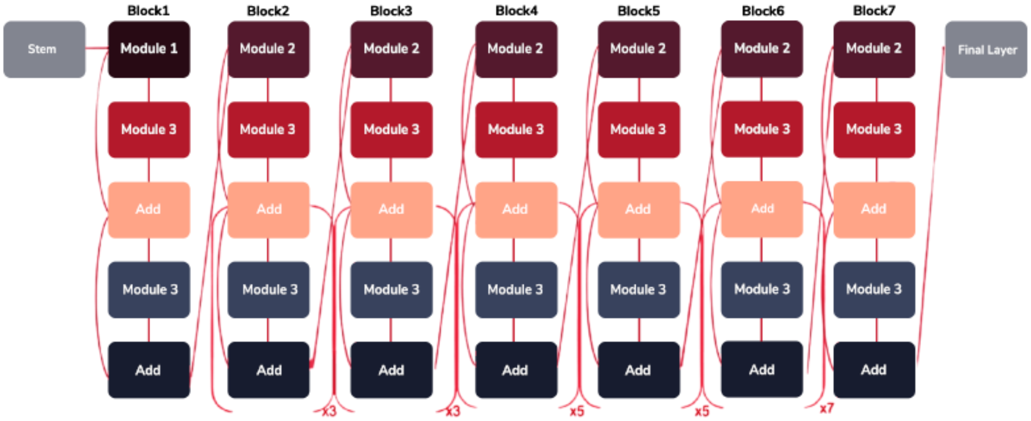


Figure 5. The architecture of EfficientNet-B5.

3.3.3. Hyper Model Fusion and Design

The proposed hyper model aims to combine the strengths of both DenseNet-161 and EfficientNet B5 architectures to enhance the accuracy of baggage detection and classification. The fusion of these two architectures is achieved through the use of a weighted average method, where the outputs of both models are combined based on their respective weights. These weights are learned during the training process to optimize the overall performance of the hyper model.

The decision to fuse DenseNet-161 and EfficientNet-B5 into a hyper model was guided by their respective strengths. DenseNet-161, with its dense connections excels at feature extraction, mitigating the vanishing-gradient problem that often affects the performance of deep networks. On the other hand, EfficientNet-B5, built on the principle of compound scaling, offers superior performance with a lower computational cost. The fusion of these models through the weighted average method empowers the hyper model to leverage the profound feature extraction capabilities of DenseNet-161 and the efficient resource utilization of EfficientNet-B5. Furthermore, these weights for fusion are not pre-set but are adaptively learned during the training, allowing the model to optimize its reliance on each base model for each prediction.

To design the hyper model, the outputs of the DenseNet-161 and EfficientNet B5 models are first concatenated before being passed through a fully connected layer. This layer is then followed by a SoftMax activation function, which produces the final class of probabilities for the five baggage classes. The hyper model is trained end-to-end using the fit-one-cycle method, as described in the subsequent sub-section. This sophisticated architecture ensures a cohesive system that effectively combines two different architectures to deliver superior performance in the complex task of baggage classification.

3.3.4. The Fit-One-Cycle Method

The fit-one-cycle method [36,37] is an innovative training approach that aims to decrease the training time and enhance model accuracy. This method involves varying the learning rate and momentum during training using a cyclical schedule. Specifically, the learning rate is gradually increased from a minimum value to a maximum value and then decreased back to the minimum, while the momentum follows the opposite pattern. By employing the fit-one-cycle method, our model can adaptively adjust its learning rate and momentum during training, leading to faster convergence and improved accuracy.

3.3.5. Evaluation Criteria

The evaluation criteria used for assessing the classification performance of our DenseNet-161 and EfficientNet B5 models, as well as the hyper model, include several metrics, such as accuracy, macro-F1, and micro-F1.

As depicted in Equation (1), accuracy measures the proportion of correctly classified instances among the total instances. It takes into account the true positives (TPs) and true negatives (TNs), which represent the number of correctly identified positive instances (baggage correctly classified as positive) and negative instances (non-baggage correctly classified as negative), respectively. Equation (2) defines the F1-score as the harmonic average between precision and recall. Precision, calculated as the TP divided by the sum of TPs and false positives (FPs), quantifies the accuracy of positive predictions. Recall, calculated as the TP divided by the sum of TPs and false negatives (FNs), measures the proportion of positive instances that are correctly identified. In Equation (3), macro-F1 calculates the average F1-score for each class, assigning equal weight to each class. In Equation (4), micro-F1 aggregates the contributions of all classes to determine the average F1-score, emphasizing global performance.

By considering the metrics of accuracy, macro-F1, and micro-F1, we can comprehensively evaluate the effectiveness of the proposed hyper model in detecting and classifying human-carried baggage based on the bag type.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{2}$$

$$\text{Macro-F1} = \frac{1}{N} \times \sum F1_i \; for \; i = 1 \; To \; N \tag{3}$$

$$\text{Micro-Precision} = \frac{\sum TP_i}{(\sum TP_i + \sum FP_i \; for \; i = 1 \; to \; N)}$$

$$\text{Micro-Recall} = \frac{\sum TP_i}{(\sum TP_i + \sum FN_i \; for \; i = 1 \; to \; N)}$$

$$\text{Micro-F1} = 2 \times \frac{Micro - Precision \times Micro - Recall}{Micro - Precision + Micro - Recall} \tag{4}$$

4. Experimental Results and Discussion

This section presents the experimental results of the proposed hyper model for detecting and classifying human-carried baggage based on the baggage type. We trained and evaluated three models–DenseNet-161, EfficientNet B5, and the hyper model that combines both models. These models were trained on the re-annotated PETA dataset using the Fastai framework [38], which is built on top of PyTorch. All our experiments were conducted using Google Colab [39], and we specifically utilized the T4 GPU for our computations.

The following subsection presents the results obtained for each model, followed by a comparison with existing techniques and a discussion and performance analysis.

4.1. Classification Results

We conducted experiments to evaluate the performance of the proposed models on the re-annotated PETA dataset. The dataset was randomly divided into the following three sets: a training set comprising 65%, a validation set comprising 15%, and a testing set comprising 20%. A batch size of 32 was used, and the training process was run for 40 epochs, except for the hyper model, which was trained for only 10 epochs. In all the experiments, we randomly selected 2000 sample images from each class for classification. Through extensive experiments, we determined that this number was optimal, adhering to the recommended rule of thumb of having at least 1000 images per class [40].

4.1.1. Densenet-161 Results

DenseNet-161 model performance metrics, such as precision, recall, and F1-score, were calculated for each of the five classes. The classification results are summarized in Table 2, while Table 3 displays the precision, recall, and F1-scores.

Table 2. Densenet-161 classification results.

Network	Accuracy	Macro-F1	Micro-F1
Dense-Net 161	95.5%	95.4%	95.5%

Table 3. Precision, recall, and F1-score.

	Precision	Recall	F1-Score
Carrying Backpack	0.97	0.97	0.97
Carrying Luggage Case	0.94	0.99	0.96
Carrying Messenger Bag	0.95	0.92	0.93
Carrying Nothing	0.97	0.94	0.95
Carrying Other	0.96	0.96	0.96
Average	0.958	0.956	0.954

The DenseNet-161 model achieved an overall accuracy of 95.5%. The loss curve of the model, as shown in Figure 6, indicates a steady decrease in loss during the training process with no apparent signs of overfitting.

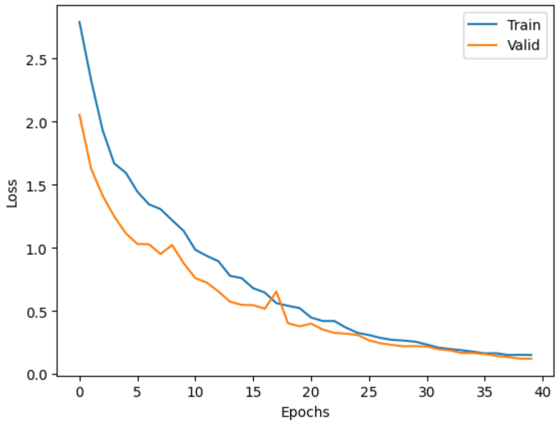


Figure 6. The loss curve for DenseNet-161 model.

The confusion matrix in Figure 7 shows the distribution of true and predicted class labels. The diagonal elements represent the number of correctly classified instances, while off-diagonal elements indicate misclassifications. From the confusion matrix, it is evident that the DenseNet-161 model performed well in classifying most instances, particularly in the Carrying Backpack and Carrying Luggage Case categories, with a high number of correctly classified instances. The model exhibited some confusion between Carrying Messenger Bags and Carrying Luggage Cases. However, overall, the model exhibited high accuracy in correctly identifying the objects that individuals are carrying.

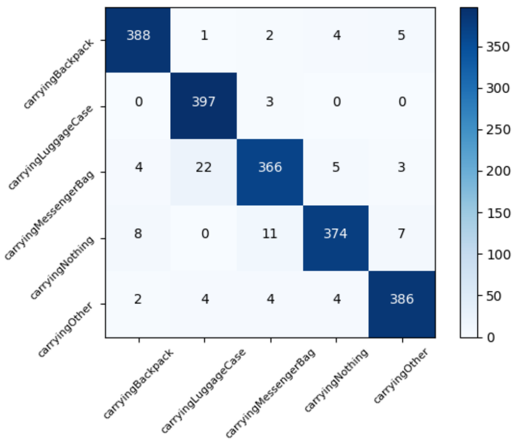


Figure 7. Confusion matrix for DenseNet-161 model.

The ROC (Receiver Operating Characteristic) curve, shown in Figure 8, demonstrates the performance of the DenseNet-161 model at various classification thresholds. Based on the provided information, it appears that this model performed very well in classifying images into different classes based on baggage carried.

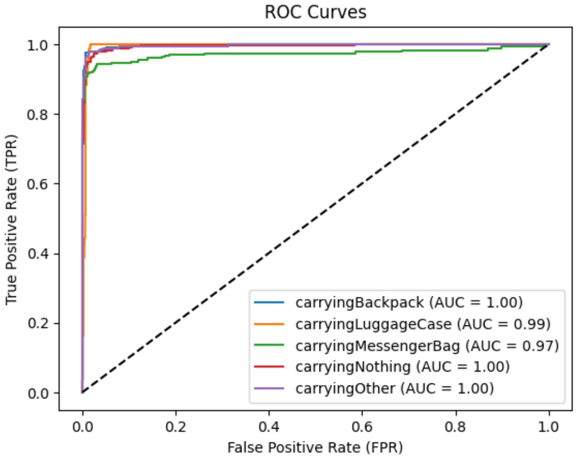


Figure 8. ROC curve for DenseNet-161 model.

In summary, the DenseNet-161 model demonstrated excellent performance in detecting and classifying human-carried baggage based on the bag type, with an overall accuracy of 95.5%. The model performed well across all five classes, as evidenced by the high precision, recall, and F1-score values. Additionally, the average time taken for all epochs was 1 min and 41 s.

4.1.2. EfficientNet B5 Results

The EfficientNet B5 model performance metrics, including precision, recall, and F1-score, were calculated for each of the five classes. The classification results are summarized in Table 4, while Table 5 displays the precision, recall, and F1-scores.

Table 4. EfficientNet B5 classification results.

Network	Accuracy	Macro-F1	Micro-F1
EfficientNet-B5	97.3%	97.2%	97.3%

Table 5. Precision, recall, and F1-score.

	Precision	Recall	F1-Score
Carrying Backpack	0.97	0.97	0.97
Carrying Luggage Case	0.97	0.98	0.98
Carrying Messenger Bag	0.98	0.95	0.96
Carrying Nothing	0.97	0.97	0.97
Carrying Other	0.98	0.99	0.98
Average	0.974	0.972	0.972

The EfficientNet B5 model achieved an overall accuracy of 97.3%. The loss curve of the model, presented in Figure 9, demonstrates the consistent decrease in loss during the training process, with no apparent signs of overfitting.

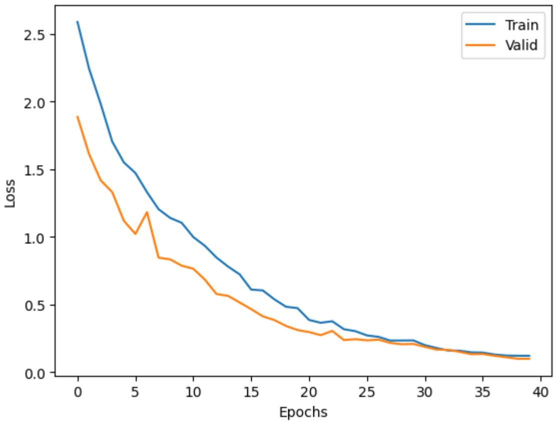


Figure 9. Loss curve for EfficientNet B5 model.

The confusion matrix depicted in Figure 10 shows the distribution of the true and predicted class labels for the classification task. The EfficientNet B5 model displayed a strong performance in accurately classifying the baggage carried by individuals, with a high number of instances correctly labeled for each class. However, this model also exhibited some confusion between Messenger Bags and Luggage Cases, which could be due to the similarity in features and the impact of data augmentation on luggage cases. Despite this, the model maintained high object identification accuracy.

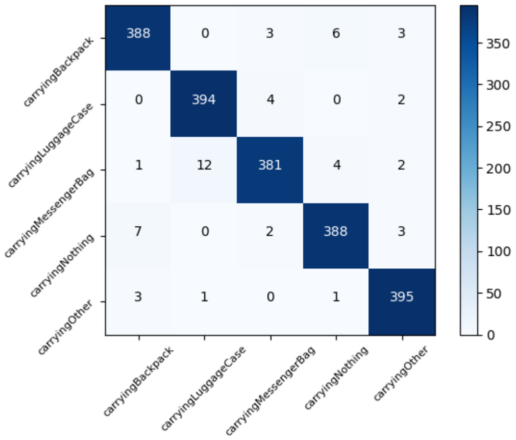


Figure 10. Confusion matrix for EfficientNet B5 model.

The ROC curve, illustrated in Figure 11, depicts the performance of the EfficientNet B5 model at various classification thresholds. The high AUC scores for each class indicate a strong classification performance. Therefore, based on the given information, it can be inferred that this model performed very well in accurately classifying images of people carrying different types of baggage into various classes.

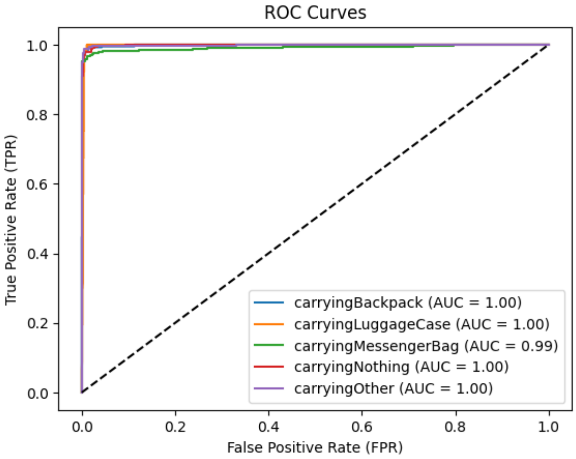


Figure 11. ROC curve for EfficientNet B5 model.

In summary, the EfficientNet B5 model exhibited exceptional performance in detecting and classifying human-carried baggage based on bag type, achieving an overall accuracy of 97.3%. The model performed well across all five classes, as evidenced by the high precision, recall, and F1-score values. Comparatively, the EfficientNet B5 model outperformed the DenseNet-161 model in terms of overall accuracy and individual class performance. Additionally, the average time taken for all epochs was 1 min and 32 s.

4.1.3. Hyper Model Results

The hyper model is an ensemble technique that combines the predictions of multiple individual models, in this case, the DenseNet-161 and EfficientNet B5 models. The model performance metrics, including precision, recall, and F1-score, were calculated for each of the five classes. The classification results are summarized in Table 6, while Table 7 displays the precision, recall, and F1-scores.

Table 6. Hyper model classification results.

Network	Accuracy	Macro-F1	Micro-F1
Hyper Model	98.65%	98.6%	98.65%

Table 7. Precision, recall, and F1-score.

	Precision	Recall	F1-Score
Carrying Backpack	0.99	0.98	0.98
Carrying Luggage Case	0.99	0.99	0.99
Carrying Messenger Bag	0.98	0.97	0.98
Carrying Nothing	0.98	0.99	0.99
Carrying Other	0.99	0.99	0.99
Average	0.986	0.984	0.986

The hyper model achieved an overall accuracy of 98.65%. This represents an improvement over the individual DenseNet-161 and EfficientNet B5 models. The ensemble technique effectively leverages the strengths of both models to achieve a better classification performance. The loss curve of the hyper model, presented in Figure 12, demonstrates a consistent decrease in loss during the training process, with no apparent signs of overfitting.

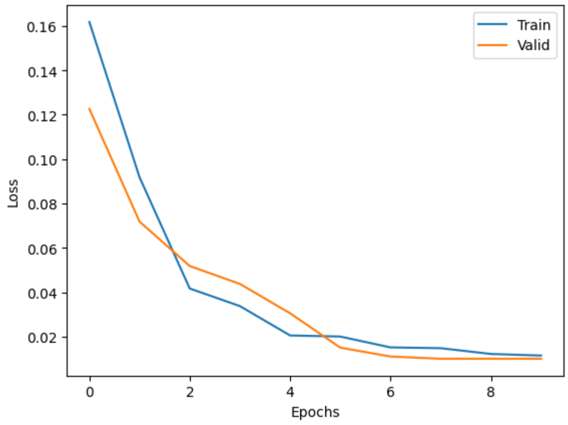


Figure 12. Loss curve for hyper model.

The confusion matrix in Figure 13 provides an overview of the true and predicted class labels' distribution for the classification task. The hyper model demonstrated an impressive performance by accurately classifying the baggage carried by individuals, with a high number of instances correctly labeled for each class. The model showed high accuracy for all classes, with only a few misclassifications, owing to the combination of strengths from DenseNet-161 and EfficientNet B5. Therefore, the confusion matrix indicates that this model is effective in accurately classifying the baggage carried by individuals, making it useful in various applications that require object recognition and classification.

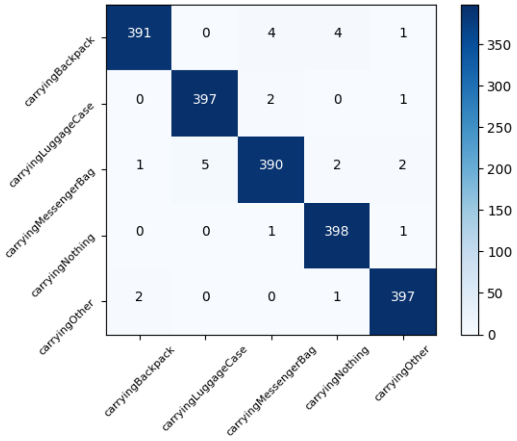


Figure 13. Confusion matrix for hyper model.

The ROC curve, depicted in Figure 14, demonstrates the performance of the hyper model at various classification thresholds. The high AUC scores observed across all classes are indicative of this model's ability to accurately distinguish between the various categories. This performance is attributed to the synergistic benefits derived from the combination of DenseNet-161 and EfficientNet B5.

In summary, the ROC curve provides compelling evidence of the hyper model's ensemble robustness and efficacy in image classification tasks involving people carrying baggage. The observed high AUC scores validate this model's exceptional performance, thereby affirming its suitability for real-world applications. Additionally, the average time taken for all epochs was 2 min and 35 s.

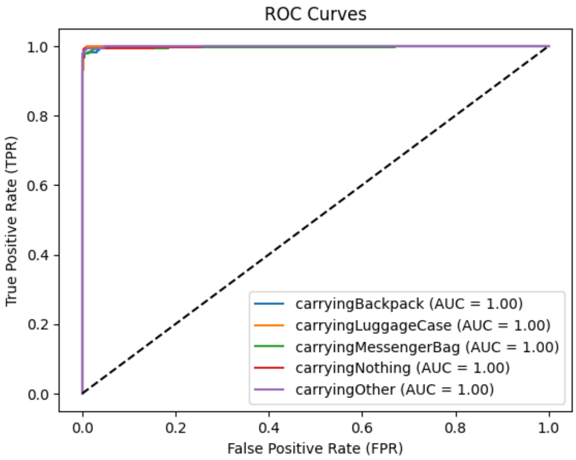


Figure 14. ROC curve for hyper model.

4.2. Comparison with Existing Techniques

In this section, we compare the performance of three models—DenseNet-161, EfficientNet B5, and the proposed Hyper Model ensemble—with existing techniques used for human-carried baggage classification. Table 8 summarizes the classification results of various methods, including the two models and the proposed hyper model ensemble developed in this study.

Table 8. Comparison of classification results with existing techniques.

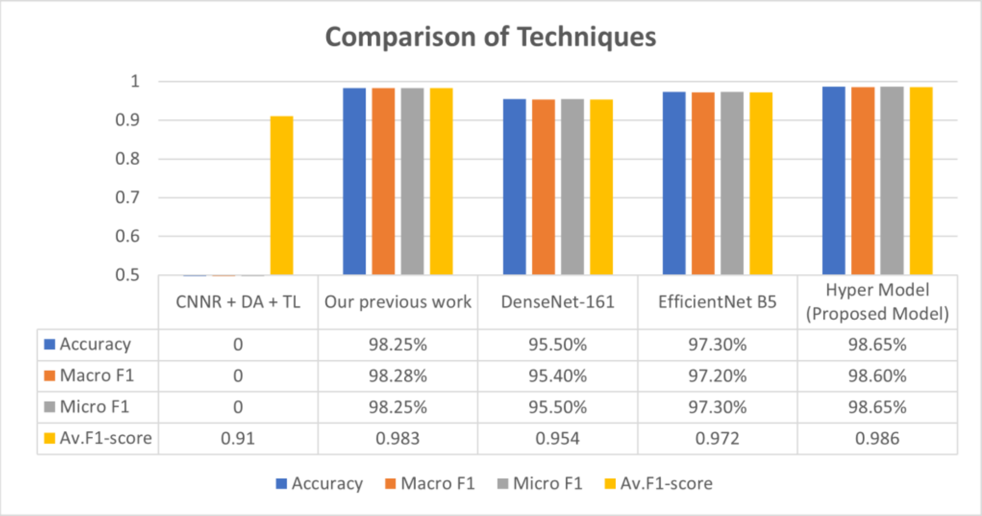
REF	Method	Accuracy	Macro-F1	Micro-F1	Av. F1-Score
[29]	CNNR + DA + TL	–	–	–	0.91
[30]	Our previous work (DenseNet-161)	98.25%	98.28%	98.25%	0.983
–	DenseNet-161	95.5%	95.4%	95.5%	0.954
–	EfficientNet B5	97.3%	97.2%	97.3%	0.972
–	Hyper Model (Proposed Model)	98.65%	98.6%	98.65%	0.986

Note: The dash symbol ("−") indicates that certain values were not available or were not addressed by the respective researchers in their publications.

As shown in Table 8, the hyper model ensemble achieved the highest overall accuracy and F1-score among all the techniques. The ensemble technique effectively combined the strengths of DenseNet-161 and EfficientNet B5 models to achieve a superior classification performance.

Figure 15 provides a visual representation that contrasts the performance metrics of DenseNet-161, EfficientNet B5, and the hyper model ensemble against established techniques, focusing on overall accuracy, macro-F1, micro-F1, and the average F1-score. The delineated results underscore the pronounced advancements that the proposed models bring forth in comparison to traditional methods.

In summary, the DenseNet-161, EfficientNet B5, and hyper model ensemble manifested an unparalleled proficiency in the detection and classification of human-carried baggage categorized by bag type. Notably, the ensemble technique warrants deeper investigation, as it showcases promise for enhancing object classification endeavors across diverse real-world applications.



	CNNR + DA + TL	Our previous work	DenseNet-161	EfficientNet B5	Hyper Model (Proposed Model)
■ Accuracy	0	98.25%	95.50%	97.30%	98.65%
■ Macro F1	0	98.28%	95.40%	97.20%	98.60%
■ Micro F1	0	98.25%	95.50%	97.30%	98.65%
■ Av.F1-score	0.91	0.983	0.954	0.972	0.986

Figure 15. Comparison of techniques—accuracy and F1-score.

4.3. Discussion and Performance Analysis

In this section, we critically evaluate the performance of DenseNet-161, EfficientNet B5, and the hyper model ensemble in the classification of human-carried baggage. We reflect on the effectiveness of these models, delve into their unique architectures, and discuss their advantages over traditional methods. Additionally, we identify limitations in our current approach and propose potential directions for future research, aiming to enhance the models' applicability and robustness in real-world scenarios.

4.3.1. Effectiveness of the Models

This study's experimental results demonstrate the effectiveness of DenseNet-161, EfficientNet B5, and the hyper model ensemble in classifying human-carried baggage based on bag types. These models outperformed existing techniques, including custom CNN (CNNR + DA + TL) and a DenseNet-161 model informed by the human viewing direction.

4.3.2. Architectural Advantages

DenseNet-161 and EfficientNet B5's standout performance can be attributed to their unique architectures and design principles. DenseNet-161, with its dense connections, mitigates the vanishing gradient problem, enhances feature propagation, and fosters feature reuse. EfficientNet B5 scales the network depth, width, and resolution harmoniously, contributing to both efficiency and accuracy.

4.3.3. Ensemble Approach

Our hyper model ensemble, which blends DenseNet-161 and EfficientNet B5, further boosts classification accuracy. This ensemble strategy, known for reducing overfitting and enhancing generalization, harnesses varied predictions from multiple models, showing significant promise in advancing object classification in practical scenarios.

4.3.4. Identified Limitations and Future Directions

1. The Dataset's Size and Diversity:
    - Current Limitation: Our study used 2000 images per class from the 19,000-image PETA dataset. While this sample size was optimal in our experiments, it has limitations.
    - Future Research: Employing larger and more diverse datasets could improve these models' generalizability and robustness. This expansion can help the models avoid overfitting to better handle varied real-world scenarios.

2. Real-Time Performance:
    - Current Limitation: The computational demands of DenseNet-161 and EfficientNet B5 pose challenges for real-time applications.
    - Future Research: Exploring model optimization techniques, such as compression, pruning, and efficient deployment on edge devices, could make these models more viable for real-time use.
3. Transfer Learning and Domain Adaptation:
    - Future Research: There is potential to adapt these models to related tasks, such as baggage content classification or anomaly detection, using transfer learning and domain adaptation techniques.
4. Multi-Object Detection:
    - Current Limitation: Our research was limited to a maximum of five classes.
    - Future Research: Expanding the model to accommodate a larger number of classes could enable comprehensive multi-class classification to a greater extent, which is crucial for broader applications.

In summary, this research validates the capability of DenseNet-161, EfficientNet B5, and the hyper model ensemble to classify objects within human-carried baggage. While these models demonstrate exceptional precision, there is room for further development, particularly in expanding their applicability to diverse object classification and detection scenarios. Future studies should focus on enhancing multi-class classification capabilities, thereby broadening the range of identifiable objects. This work lays a foundational benchmark for future advancements in sophisticated object classification and detection methods.

5. Conclusions

In this research, we propose a novel hyper model ensemble for detecting and classifying human-carried baggage based on baggage types in video surveillance systems. The Fastai framework was leveraged to combine the strengths of DenseNet-161 and EfficientNet-b5 pretrained models, resulting in enhanced accuracy and robustness. We utilized the PETA dataset, automatically re-annotating it into five classes corresponding to baggage types, and implemented an effective pre-processing pipeline to optimize the performance of our model. Furthermore, we employed the fit-one-cycle policy to expedite the training time while simultaneously improving model accuracy.

Our experimental results demonstrate the efficacy of the proposed hyper model ensemble, achieving an impressive accuracy of 98.6%. This surpasses existing methodologies in terms of accuracy, macro-F1, and micro-F1. The demonstrated potential of our hyper model ensemble in improving the detection and classification of human-carried baggage in video surveillance systems opens avenues for future research, such as integrating our model with object detection algorithms like Faster R-CNN, which could further enhance baggage detection performance.

Author Contributions: Conceptualization, M.A.S.; Data curation, M.A. and M.K.R.; Formal analysis, M.A.S. and M.K.R.; Investigation, M.K.R.; Methodology, M.K.R.; Project administration, M.A.S.; Resources, M.A.; Software, M.K.R.; Supervision, M.K.R.; Validation, M.A.S. and M.A.; Visualization, M.A.; Writing—original draft, M.K.R.; Writing—review and editing, M.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the reason that we used the public datasets, and the figures with humans are from public datasets.

Informed Consent Statement: Informed consent was waived for this study due to the reason that we used the public datasets and because the figures with humans are from public datasets.

Data Availability Statement: The PETA dataset used in the experiments is publicly available online at: http://mmlab.ie.cuhk.edu.hk/projects/PETA.html and accessed on 16 May 2023.

Acknowledgments: The authors are grateful to the reviewers and editors for providing advice and guidance on to how to improve the paper.

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

References

1. Elharrouss, O.; Almaadeed, N.; Al-Maadeed, S. A Review of Video Surveillance Systems. *J. Vis. Commun. Image Represent.* **2021**, *77*, 103116. [CrossRef]
2. Mishra, P.K.; Saroha, G.P. A Study on Video Surveillance System for Object Detection and Tracking. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 221–226.
3. Liu, H.; Chen, S.; Kubota, N. Intelligent Video Systems and Analytics: A Survey. *IEEE Trans. Ind. Inform.* **2013**, *9*, 1222–1233.
4. Zhao, Z.-Q.; Zheng, P.; Xu, S.; Wu, X. Object Detection with Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef]
5. Bayoumi, R.M.; Hemayed, E.E.; Ragab, M.E.; Fayek, M.B. Person Re-Identification via Pyramid Multipart Features and Multi-Attention Framework. *Big Data Cogn. Comput.* **2022**, *6*, 20. [CrossRef]
6. Jha, S.; Seo, C.; Yang, E.; Joshi, G.P. Real Time Object Detection and Trackingsystem for Video Surveillance System. *Multimed. Tools Appl.* **2021**, *80*, 3981–3996. [CrossRef]
7. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
8. Chang, L.C.; Pare, S.; Meena, M.S.; Jain, D.; Li, D.L.; Saxena, A.; Prasad, M.; Lin, C.T. An Intelligent Automatic Human Detection and Tracking System Based on Weighted Resampling Particle Filtering. *Big Data Cogn. Comput.* **2020**, *4*, 27. [CrossRef]
9. Yanagisawa, H.; Yamashita, T.; Watanabe, H. A Study on Object Detection Method from Manga Images Using CNN. In Proceedings of the 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 7–9 January 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–4.
10. Wang, N.; Wang, Y.; Er, M.J. Review on Deep Learning Techniques for Marine Object Recognition: Architectures and Algorithms. *Control Eng. Pract.* **2022**, *118*, 104458. [CrossRef]
11. Khanam, T.; Deb, K. Baggage Recognition in Occluded Environment Using Boosting Technique. *KSII Trans. Internet Inf. Syst.* **2017**, *11*, 5436–5458.
12. Han, W.; Chen, J.; Wang, L.; Feng, R.; Li, F.; Wu, L.; Tian, T.; Yan, J. Methods for Small, Weak Object Detection in Optical High-Resolution Remote Sensing Images: A Survey of Advances and Challenges. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 8–34. [CrossRef]
13. Xu, J. A Deep Learning Approach to Building an Intelligent Video Surveillance System. *Multimed. Tools Appl.* **2021**, *80*, 5495–5515. [CrossRef]
14. Gouiaa, R.; Akhloufi, M.A.; Shahbazi, M. Advances in Convolution Neural Networks Based Crowd Counting and Density Estimation. *Big Data Cogn. Comput.* **2021**, *5*, 50. [CrossRef]
15. Popoola, O.P.; Wang, K. Video-Based Abnormal Human Behavior Recognition—A Review. *IEEE Trans. Syst. Man Cybern. Part C* **2012**, *42*, 865–878. [CrossRef]
16. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 221–231. [CrossRef]
17. Baccouche, M.; Mamalet, F.; Wolf, C.; Garcia, C.; Baskurt, A. Sequential Deep Learning for Human Action Recognition. In Proceedings of the Human Behavior Understanding: Second International Workshop, HBU 2011, Amsterdam, The Netherlands, 16 November 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 29–39.
18. Donahue, J.; Anne Hendricks, L.; Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Saenko, K.; Darrell, T. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2625–2634.
19. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
20. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015; Volume 28.
21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single Shot Multibox Detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
22. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
23. Dewi, C.; Chen, A.P.S.; Christanto, H.J. Deep Learning for Highly Accurate Hand Recognition Based on Yolov7 Model. *Big Data Cogn. Comput.* **2023**, *7*, 53. [CrossRef]

24. Ahammed, M.T.; Ghosh, S.; Ashik, M.A.R. Human and Object Detection Using Machine Learning Algorithm. In Proceedings of the 2022 Trends in Electrical, Electronics, Computer Engineering Conference (TEECCON), Bengaluru, India, 26–27 May 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 39–44.
25. Chen, L.; Zhang, H.; Xiao, J.; Nie, L.; Shao, J.; Liu, W.; Chua, T.-S. Sca-Cnn: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5659–5667.
26. Fu, J.; Zheng, H.; Mei, T. Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4438–4446.
27. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
28. Sun, C.; Shrivastava, A.; Singh, S.; Gupta, A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In Proceedings of the Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 843–852.
29. Jo, K.-H. Human Carrying Baggage Classification Using Transfer Learning on CNN with Direction Attribute. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Intelligent Computing, Liverpool, UK, 7–10 August 2017*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 717–724.
30. Ramadan, M.K.; Youssif, A.A.A.; El-Behaidy, W.H. Detection and Classification of Human-Carrying Baggage Using DenseNet-161 and Fit One Cycle. *Big Data Cogn. Comput.* **2022**, *6*, 108. [CrossRef]
31. Deng, Y.; Luo, P.; Loy, C.C.; Tang, X. Pedestrian Attribute Recognition at Far Distance. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 7 November 2014; pp. 789–792.
32. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
33. Krig, S.; Krig, S. Image Pre-Processing. In *Computer Vision Metrics (Textbook Edition)*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 35–74.
34. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
35. Tan, M.; Le, Q. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; PMLR: Westminster, London, 2019; pp. 6105–6114.
36. Smith, L.N. Cyclical Learning Rates for Training Neural Networks. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 464–472.
37. Smith, L.N.; Topin, N. Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates. In Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, Baltimore, MD, USA, 15–17 April 2019; SPIE: Bellingham, WA, USA, 2019; Volume 11006, pp. 369–386.
38. Howard, J.; Gugger, S. Fastai: A Layered API for Deep Learning. *Information* **2020**, *11*, 108. [CrossRef]
39. Bisong, E.; Bisong, E. Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*; Apress: Berkeley, CA, USA, 2019; pp. 59–64.
40. Shahinfar, S.; Meek, P.; Falzon, G. "How Many Images Do I Need?" Understanding How Sample Size per Class Affects Deep Learning Model Performance Metrics for Balanced Designs in Autonomous Wildlife Monitoring. *Ecol. Inform.* **2020**, *57*, 101085. [CrossRef]

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