



of identification that involve the use of ink or heat to create a permanent mark on an animal's skin [1,2]. However, these techniques can be painful and are prone to alteration or copying onto other animals. Microchips injected into the nuchal ligament [3] and containing a serial number can also be used for identification. According to recent studies, microchip implantation in animals has generally been considered a safe and effective means of identification. However, as with any medical procedure, there are potential risks and limitations to be considered.

One potential risk of microchip implantation is the possibility of infection at the site of the implant. Other potential risks include migration of the microchip within the animal's body, malfunction of the microchip, or adverse reactions to the microchip. These risks are generally considered to be rare, but it is important for animal owners to be aware of them and to discuss any concerns with their veterinarian.

In addition to the potential risks, there are also some drawbacks to consider when using microchips for animal identification. One potential disadvantage is the cost of the microchip and the implantation procedure, which may be a barrier for some animal owners. Additionally, microchips can only be read with specialized scanners, which may not always be readily available. This means that if an animal becomes lost, it may not be immediately identifiable if the finder does not have access to a scanner. In addition, chip scanners cannot read all the chips [4]. Finally, microchips do not provide any visible identification for animals, which can make it difficult to identify them if the microchip is not scanned. Moreover, if the chips are made smaller so that they stay in one place and do not cause any damage, their readability is affected [5].

In addition to these traditional methods, biometric identification techniques, such as iris matching, DNA matching, and muzzle pattern matching, offer unique and reliable means of identification [6]. Iris feature scanning uses mathematical techniques to identify patterns in the visual image of the eye, while retinal imaging utilizes the unique vascular pattern of an animal's retina. External patterns, such as muzzle patterns, can also be photographed or filmed for identification. Digital images can be manipulated to improve recognition accuracy.

Livestock biometric identification involves the use of unique physical characteristics, such as facial features, hoof prints, and retina patterns, to accurately identify individual animals. Muzzle patterns, which are the unique patterns of fur and skin on the muzzle of an animal, have been found a promising biometric feature for identification in recent studies. Muzzle patterns are unique to each animal and can serve as a reliable means of identification [7]. The use of muzzle patterns for livestock identification has the potential to improve various aspects of modern livestock industries.

One key benefit of using muzzle patterns for livestock identification is improved traceability. Accurate identification of animals enables the tracking of their movement and history, which is essential for food safety and disease control. In the event of an outbreak of a foodborne illness or a disease that spreads among livestock, accurate identification can help to quickly trace the source and contain the spread.

The use of muzzle patterns for livestock identification can also improve breeding programs by enabling the accurate identification and tracking of breeding stock. This can help to improve the quality and efficiency of breeding programs, as well as the overall genetic diversity and health of the livestock population.

In addition to traceability and breeding programs, accurate livestock identification using muzzle patterns can improve the overall management and organization of livestock. It can help to ensure that animals are properly cared for, housed, and managed, as well as to track their production efficiency and welfare.

This study proposes a system using the biometrics of livestock animals, such as their muzzle patterns, for the automated, instant identification of livestock, utilizing deep transfer learning to detect the face and the muzzle point of the animal. The system comprises three steps that include the detection of the animal's face in the image to ensure that the animal is present, and the face is again processed to detect the nose or muzzle point of

the animal; both of these detections are achieved using the state-of-the-art you only look once (Yolo v7) [8] object detection model. Any noise or image with motion blur is filtered out in consideration of any movement by the animal, and only clear images of the nose are captured, which are then used to extract features using SIFT (scale-invariant feature transform) [9] and then stored in the database. Later on, the initial steps are the same when recognizing but, after the features are extracted, the system uses an approximate nearest neighbor search by applying FLANN (fast library for approximate nearest neighbor) matcher [10,11] to match SIFT-extracted features.

The THDD dataset [12,13] was applied to train the Yolo object detector in this study as an example of the proposed approach, due to the lack of available data for all livestock animals. The unique features of livestock animals such as cattle, goats, sheep, and horses/equines in their noses or muzzles can be used for identification purposes [2] by assigning an ID and other relevant information to their extracted features and retrieving any associated information upon recognition.

In summary, the use of muzzle patterns for livestock identification has the potential to improve traceability, disease control, breeding programs, and the overall management and organization of livestock. It is also important for the monitoring of animal welfare and production efficiency. The use of biometric identification technologies, such as image scanning and automated detection and feature extraction algorithms, can enhance the accuracy and efficiency of this process. Accurate identification of livestock is essential for the success of modern livestock industries and the proper functioning of related systems. The demand for various pattern recognition applications has drastically increased during the last couple of decades. Considering the importance and benefits of biometric identification, livestock farmers are more eager to deploy the most efficient and cost-effective system.

## 2. Related Work

This literature review aims to examine the current state of research on the use of biometric feature extraction and recognition for livestock face recognition, including the benefits, challenges, and future directions of this approach.

Livestock face recognition using biometric feature extraction and recognition techniques has gained significant attention in recent years as a means of improving the traceability, disease control, breeding programs, and overall management and organization of livestock. The use of unique physical characteristics, such as muzzle patterns, for the accurate identification of animals, has shown promising results in several studies. In this context, the use of biometric identification technologies, such as automated detection and feature extraction algorithms, can enhance the accuracy and efficiency of livestock identification.

Taha et al. [14] proposed a technique for Arabian horse identification based on the muzzle patterns of the horses. They used a fusion of local binary pattern (LBP) and speed-up-robust-features (SURF) techniques to extract features from horses' nose images. The extracted features were then used for classification using a support vector machine (SVM) classifier, which achieved accuracy of 96%. The authors also applied optimization using a gray-wolf algorithm, which further increased the accuracy to 99%.

This study demonstrates the use of muzzle patterns for Arabian horse identification. However, the study also has some limitations. One potential limitation is the issue of scale invariance, which refers to the ability of the identification system to accurately recognize the animal regardless of the size of the image. Some identification techniques, such as LBP and SURF, may not be scale-invariant and are sensitive to changes in the size of the image. This can be a problem when dealing with real-world images, as the size of the animal and the distance of the camera from the animal can vary significantly.

To address the issue of scale invariance, researchers have explored the use of techniques such as SIFT [9] or histogram of oriented gradients (HOG), which are designed to be more robust to changes in the scale of the image. Other potential approaches for improving the robustness of animal identification systems include the use of ensemble methods or deep

learning techniques, which are effective in a variety of image recognition tasks. In addition, it is important to carefully consider the size and diversity of the training dataset and to optimize the various components of the identification system to ensure the best possible performance.

Kumar et al. [15] proposed a biometric identification technique for cattle using their muzzle prints. The authors extracted the features of muzzle patterns using Gaussian pyramid levels, with different approaches such as appearance-based feature extraction, texture descriptor-based feature extraction, and representation. They used the technique of chi-square-based matching to match the extracted features for the identification of the animal. The proposed algorithm achieved accuracy of 93.87% for the identification of cattle using their muzzle patterns.

This study highlights the potential of using muzzle patterns for cattle identification and demonstrates the effectiveness of the proposed feature extraction and matching techniques. However, the study also has some limitations. The authors used live-captured muzzle images of the cattle face, taking the picture-capturing device too close to the face of the animal, which can be a problem because the animals may become agitated or scared at the sight of a camera or phone and may not be stable enough for the captured image to obtain a clear muzzle.

The use of live-captured images for animal identification can be challenging due to the inherent variability in the quality of the images and the potential for the animal to become distressed or agitated. To overcome these challenges, the proposed system uses alternative methods of automating the process for obtaining images; this method can allow for the capture of images from a distance, reducing the need for close contact with the animal and minimizing the risk of distress.

In addition to the challenges of image capture, other factors can impact the accuracy of animal identification systems. These include the complexity and uniqueness of the biometric features used, the quality of the feature extraction and matching algorithms, and the size of the database of reference images. To improve the accuracy and robustness of animal identification systems, it is important to consider all of these factors and to carefully optimize the various components of the system.

In conclusion, the study by Kumar et al. [15] demonstrates the potential of using muzzle patterns for cattle identification and highlights the importance of considering the quality of the images used for identification.

Li et al. [16] proposed a cattle identification approach based on deep learning techniques. The authors used 59 different deep learning models to identify individual cattle using their muzzle prints and achieved accuracy of 98.7%. This study demonstrates the potential of deep learning for cattle identification and highlights the effectiveness of using muzzle prints as a biometric feature for identification.

However, it is challenging to use still muzzle point images for identification when the animal is moving. In such cases, it can be difficult to capture a clear image of the muzzle, which can hinder the performance of the deep learning model. This is a common limitation in animal identification, as animals tend to move frequently and may not always be in a stable position.

In conclusion, the study by Li et al. [16] demonstrates the potential of deep learning for cattle identification but it has a limitation in identifying moving animals because, if the animal is moving, the image may be distorted and lose unique information required for the model to identify the animal.

Ouarda et al. [17] proposed a face recognition system for horses using the Gabor filter and linear discriminant analysis (LDA) for facial texture features. The proposed system achieved identification accuracy of 95.74%. However, using horse faces for identification may not be practical due to the difficulty in capturing stable images of the horse's face. Horses tend to move their heads frequently, which can make it difficult to obtain clear images with sufficient features for LDA to make accurate predictions. The use of computer

vision for animal identification should prioritize the comfort and well-being of the animal, and a system that requires the animal to remain still may not be ideal.

This study highlights the importance of considering the behavior and characteristics of the animal when developing an identification system. While the proposed system by Ouarda et al. demonstrated high accuracy, its practicality may be limited due to the difficulty in obtaining stable images of the horse's face. Future research could explore alternative methods for identifying horses that do not require the animal to remain still and are more suitable for use in a natural setting.

Kumar et al. [18] proposed a real-time system for recognizing cattle using the distinctive patterns found on their muzzles or noses. The system involves capturing video footage of the cattle, cropping the frames to isolate the nose/muzzle region, preprocessing the images to remove any noise or blurriness, converting them to grayscale, extracting features using appearance-based algorithms, and storing the features in a database. During testing, a query image of a nose is taken and its features are extracted using a fisher locality preserving projections (FLLP) extractor, which is then matched against the features in the database.

A threshold value is returned, which is compared to a predetermined threshold value to determine whether the recognition of the cattle is successful. However, one limitation of this method is that the nose images are derived from surveillance video frames, rather than being directly captured. This means that the accuracy of the recognition may be affected if the quality of the video or the visibility of the nose is poor.

Jarraya et al. [19] proposed a technique for horse identification using face biometrics. In their study, the authors proposed horse detection and recognition by first detecting the horse's face using a sparse neural network (SNN), which achieved average precision ranging from 84% to 90%. They used Gabor features, linear discriminant analysis (LDA), and support vector machine (SVM) for face identification, achieving accuracy of 99.89%. The authors cropped faces using MATLAB, but their model was unable to detect small faces in the images.

This study also demonstrates the potential of using face biometrics for horse identification. It highlights the effectiveness of the proposed automated feature extraction system, but the study has a limitation in that the system is unable to detect small faces in the images. This can be a problem when dealing with real-world images, as horses of different sizes and breeds may have faces of varying sizes, or it may be challenging to capture a clear image in close proximity to the animal.

In contrast, the proposed system in this study can detect both small and large faces of animals, making it more robust and suitable for use in real-world scenarios. This is achieved through the use of state-of-the-art Yolo (v7) object detection, which is effective in detecting small objects in images. In addition, the proposed system carefully considers the size and diversity of the training dataset and optimizes the various components of the identification system to ensure the best possible performance.

The limitations discussed in the literature review, such as the difficulty in capturing stable images of a moving animal or the sensitivity of certain identification techniques to changes in image size, have been addressed in the proposed system. A summary of the previous related work along with its implications is presented in Table 1.

**Table 1.** Comparison and limitations of previous approaches.

Study	Methodology	Accuracy %	Limitations
Arabian Horse Identification and Gender Determination System Based on Feature Fusion and Gray Wolf Optimization [14]	Fusion of LBP and SURF, SVM with gray-wolf optimization	99.00	Sensitive to changes in the size of the image, leading to issues with scale invariance.
Muzzle Point Pattern Based Techniques for Individual Cattle Identification [15]	Gaussian pyramid with chi-square-based matching	93.87	Requires close image capturing.
Individual Beef Cattle Identification Using Muzzle Images and Deep Learning Techniques [16]	Deep learning-based multiple models	98.70	Moving object may deteriorate accuracy.
A Preliminary Investigation on Horses Recognition Using Facial Texture Features [17]	Gabor filter with LDA	95.74	Moving object may deteriorate accuracy.
Real-time recognition of cattle using animal biometrics [18]	Fisher locality preserving projection-based feature extraction	96.87	The accuracy of recognition may be affected if the quality of the video or the visibility of the nose is poor.
Sparse Neural Network for Horse Face Detection in a Smart Riding Club Biometric System [19]	Sparse neural network, LDA, and SVM	99.89	The system is unable to detect small faces in the images.

### Limitations of Related Work

Table 1 depicts the weaknesses of existing classification approaches. Previous methods have at least one of the following limitations:

- One potential limitation of using LBP and SURF for animal identification is their sensitivity to changes in the size of the image, leading to issues with scale invariance.
- The proposed system using the Gaussian pyramid with chi-square-based matching technique requires one to be very close to the animal to capture the animal's face, which can cause agitation and fear in the animal.
- The movement of the animal can hinder the performance of the DL model and limit its ability to identify because the movement of the animal will cause motion blur in the images.
- The use of the Gabor filter and LDA for horse face recognition may not be practical due to the difficulty in capturing stable images of the horse's face, which can hinder the performance of the identification system and may not be suitable for use in a natural setting.
- The nose images are derived from surveillance video frames, rather than being directly captured. This means that the accuracy of the recognition may be affected if the quality of the video or the visibility of the nose is poor.
- The sparse neural network is unable to detect small faces in the images. This can be a problem when dealing with real-world images.

In contrast to the limitations highlighted above, our proposed model presents an innovative approach to identify the animals more accurately. Our contribution to this study is precisely described as below:

- The proposed system can detect and process both small and large faces and the SIFT algorithm is scale-invariant and thus can extract the features of small faces.
- The system uses real-time data for feature extraction and recognition.
- The proposed system is trained to detect only clear images; in the case of motion blur, the model will not detect the muzzle so that only clear images are obtained.
- The system automates the task of capturing the face of the animal by detecting it from a safe distance.
- The system uses SIFT for feature extraction, which is scale-invariant.

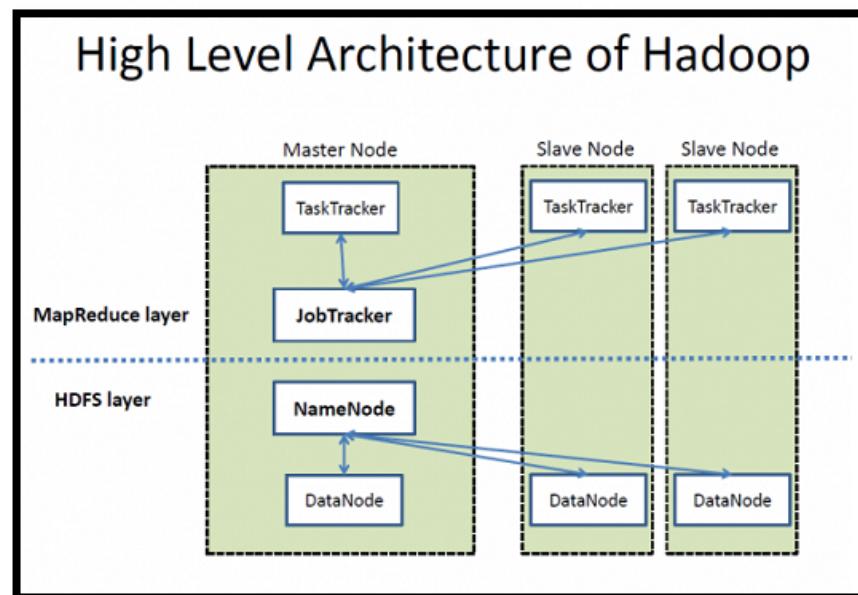
### 3. Methodology

The proposed system, DTL-AFI-VHA (Deep Transfer Learning-Based Animal Face Identification Model Empowered with a Vision-Based Hybrid Approach), utilizes advanced image processing techniques to automate the registration and identification process for livestock animals. The system is designed to detect the face and nose/muzzle of the animal using state-of-the-art object detection algorithms, such as Yolo (v7). Once the face and nose/muzzle have been detected, the system extracts distinctive features from the muzzle using the SIFT algorithm. These extracted features are then stored in a database along with the ID of the animal.

When recognition is required, the system follows the same process to detect the face and nose/muzzle of the animal and extract features using SIFT. These extracted features are then compared to the features stored in the database using FLANN algorithm. If a match is found, the system is able to accurately identify the animal with a high degree of accuracy.

The proposed system has several advantages over traditional approaches for livestock identification. One key advantage is the ability to automate the registration and identification process, reducing the need for manual intervention and improving the efficiency of the system. Additionally, the use of advanced image processing techniques, such as SIFT and FLANN, enables the system to handle a wide range of real-world scenarios, including the extraction of features from small faces.

To handle the large amounts of data generated by the system, the proposed system utilizes Apache Hadoop, an open-source framework for the distributed processing and fault tolerance of big data. The data are stored in a distributed manner on slave nodes using the Hadoop Distributed File System (HDFS)—see Figure 1—and processed in parallel using Hadoop and Apache Spark Clusters, PyTorch (a Python Library), and the Python programming language. This enables the system to efficiently handle large amounts of data and ensure the reliability of the system.

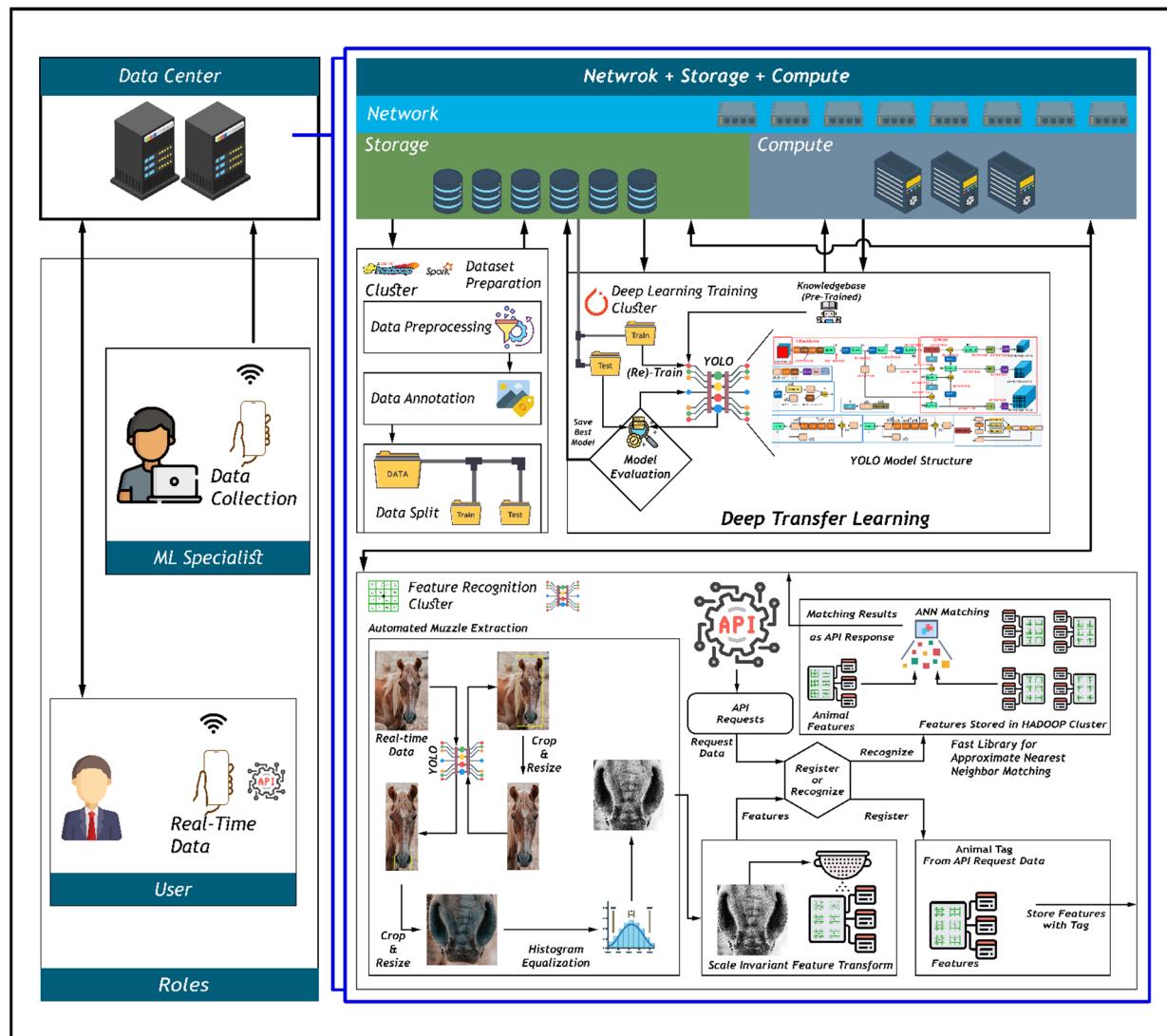


**Figure 1.** High-Level Architecture of Hadoop.

Overall, the proposed system, DTL-AFI-VHA, offers a number of benefits for the automated registration and identification of livestock animals. Its ability to detect and extract features from the faces and muzzles of animals, combined with the use of advanced image processing techniques and distributed processing on the cloud, make it a promising approach addressing the challenges of livestock identification in modern industries.

### 3.1. System Model

The proposed system, shown in Figure 2, is organized into three main modules: training an object detection model, registering the animal, and identifying the animal. The first module involves the development of a model to detect the faces and noses of animals, which is necessary in automating the extraction of the animal's nose. The second module involves the registration of the animal in the system through the use of a unique tag or serial number. The third module is responsible for the recognition or identification of the animal.



**Figure 2.** Architecture of Proposed System.

The first module of the system involves training an object detection model to detect the faces and noses of animals. This is achieved by processing unstructured data, which includes annotating and splitting the data into training and testing sets. The data are then fed to a Yolo (You Only Look Once) model, which is a type of convolutional neural network (CNN) designed for object detection. The Yolo model uses pre-trained weights, which are a set of parameters that have been previously trained on a large dataset, and bounding boxes to perform deep transfer learning. Deep transfer learning is a method of training a model on a new dataset by using the knowledge gained from a similar dataset. In this case, the Yolo model is able to use its knowledge of bounding boxes to detect the faces and noses of animals in the new dataset.

The second module of the system is responsible for registering animals in the system by associating them with a unique tag or serial number. This can be performed in real time or using a prerecorded video or image of the animal. A Rest-API (Application Programming Interface) using FastAPI is developed for this purpose, and an API request with the animal's tag is posted to register the animal. The API takes a video or image of the animal and processes it through the Yolo object detector to detect the face and nose of the animal. Once the face and nose are detected, the nose image is extracted and processed using histogram equalization and SIFT to extract keypoints and descriptors. These keypoints and descriptors are used to describe the unique features of the animal's nose, and they are stored in a cloud storage system along with the animal's identity information.

The third module of the system is used for the identification of the animal. When an API request is made with an image or real-time video stream of the animal, the nose of the animal is extracted in the same way as in the second module. The extracted features and descriptors are then matched using approximate nearest neighbor search with the FLANN algorithm. FLANN is a high-performance algorithm that utilizes two algorithms—the randomized kd-tree algorithm and the hierarchical k-means tree algorithm—to perform approximate nearest neighbor matching. If the animal has been registered in the system and a match is found, the animal's identity information is returned. If the match is not successful, or if the animal is not registered, the system will prompt the user to register the animal. The use of a threshold distance and a minimum number of good points reduces the likelihood of incorrect matches, but there is still the possibility of misidentification or false positives.

---

#### Algorithm 1: Animal Identification and Registration

---

**Input:** TAG = animal TAG;  
stream = video stream of animal / image of animal  
**Output:** Registered animal, animal's identity information  
Extracted muzzle point of animal stored in database with TAG  
**while** stream do:  
    validate(TAG)  
    frame ← stream.frame  
    frame ← frame.reshape  
    face ← detect(frame, "Face")  
    nose ← detect(face, "Nose")  
    nose\_feature ← extract\_feature (nose)  
    if registering:  
        store\_in\_database(nose\_feature, TAG, identity\_info)  
        return image, identity\_info  
    else:  
        features, descriptors, identity\_info ← retrieve\_from\_database()  
        matches ← match\_features(nose\_feature, features, descriptors)  
        if matches:  
            return identity\_info[matches [0]]  
        else:  
            return "No match found. Please register animal."  
        display(image)  
**end**

---

The validate () function is used to verify the validity of the animal tag, and the detect() and extract feature() functions are used to detect and extract the feature of the animal's nose from the video stream or image. The store in database() function is used to store the extracted nose feature in the database along with the animal tag, and the display() function is used to display the image of the extracted nose feature. The algorithm continues to run in a loop as long as the video stream is active as shown in Algorithm 1.

### 3.2. Yolo v7 Object Detector

For the detection of the faces and noses of livestock animals, the state-of-the-art object detection model Yolo v7 [8] is used. For the training of the object detector, transfer learning is used, where pre-trained or pre-learned knowledge of bounding box prediction is transferred during training. Images from the THDD horse dataset were used for the training. The total number of images in this dataset is 1103. The dataset was split randomly into training and test sets; 1000 images were used for training and 103 images for testing. The detection classes used were (1) Face, (2) Nose, (3) Nose\_ (if there are occlusions or the animal has dirt or any other debris on the nose), (4) Neg (for human faces, as humans also have the same facial features, such as two eyes, nose, mouth, and head). The total annotations were 2352, which included 1179 Face annotations, 80 for Neg, 754 Nose annotations, and 338 for Nose\_. The difference in annotations is due to the fact that some of the horses in the images were too far away and, because of this, the nose area was not annotated; see Figure 3.



**Figure 3.** An image in the data that contains an animal that is far from the camera.

The number of annotations in the training set is 2110, having 1056 annotations for the Face class, 68 for Neg, 655 for the Nose class, and 330 for the Nose\_ class, while the test set has 242 annotations, including 123 annotations for the Face class, 12 for the Neg class, 99 for the Nose class, and eight for the Nose\_ class. For annotation and data preparation, Hadoop and Apache Spark were utilized, and a data processing cluster was used for the preparation of the dataset.

Once the dataset was prepared, it was exported to the deep learning cluster, where transfer learning was applied and a new model was trained to detect animal faces and noses.

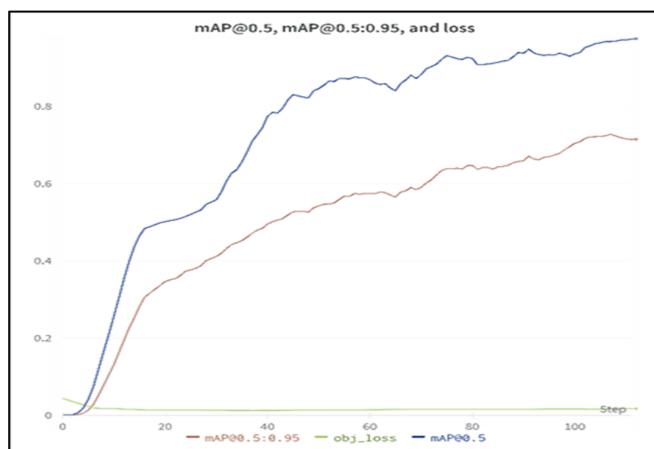
#### 3.2.1. Results and Validation of Yolo

Yolo object detector training via transfer learning resulted in 96.2% mean average precision with 0.50 as the confidence threshold for the detection of all the classes while showing 99.5%, 99.7%, and 99.5% mean average precision for the Face, the Nose, and the Nose\_ classes, respectively, while it achieved 86.3% for the Neg class, as shown in Table 2, which was expected as there were not sufficient data on this class for the model to achieve better accuracy. Nonetheless, 86.3% mean average precision with this small amount of data shows how well the Yolo objector can perform in object detection, which was the reason that we chose to use this object detection model.

**Table 2.** Metrics of Transfer Learning for the Evaluation of Model Performance.

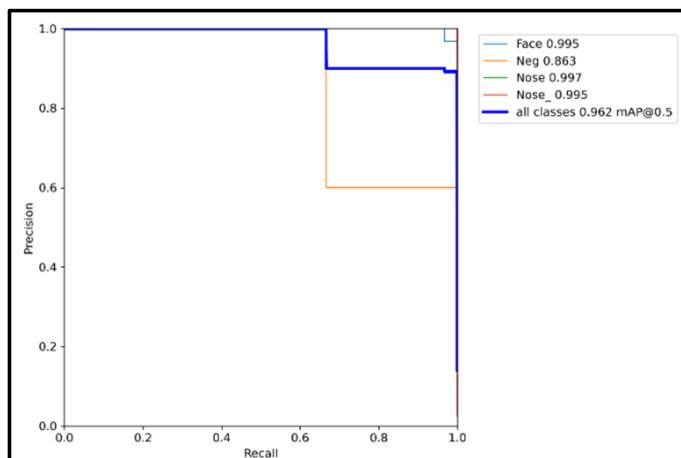
Class	Precision %	Recall %	mAP@0.5 %	mAP@0.5:0.95 %
All	99.2	89.7	96.2	74.8
Face	100.0	96.8	99.5	88.0
Neg	98.3	66.7	86.3	51.9
Nose	100.0	95.2	99.7	85.0
Nose_	98.6	100.0	99.5	74.6

The transfer learning of the model was plotted for the Yolo object detector model, which can be seen in the following Figure 4. The model was able to achieve this level of accuracy in slightly over 100 epochs. Yolo is considered a state-of-the-art object detection model at present. It can detect small objects with great accuracy and performance. Its inference time is 0.7 s on the CPU.



**Figure 4.** Mean Average Precision and Average Loss of Transfer Learning of Yolo with respect to Training Epochs.

Figure 5 shows how well the model is trained, and it also shows how well the model performs in terms of detection. The model can detect small objects or objects that are far away from the camera, which can be seen in Figure 6, where (a) contains the labels or annotations of the images from the validation set, while (b) contains the predicted bounding boxes, predicted by the Yolo object detection model with confidence. It can be seen in (b) that the Yolo model could detect the Nose\_ class, even though it was not labeled and was far away from the camera, with reasonable confidence.



**Figure 5.** Precision-Recall Curve of Transfer Learning.



**Figure 6.** Testing Object Detection Model: in (a) are the annotated images of the test set, and in (b) are the bounding boxes predicted by the Yolo object detection model trained by transfer learning.

### 3.2.2. Confusion Matrix

The confusion matrix for the validation set is shown in Table 3. The class Neg has background false negative and background false positive prediction by the model; this is due to the fact that there were not sufficient data on the Neg class.

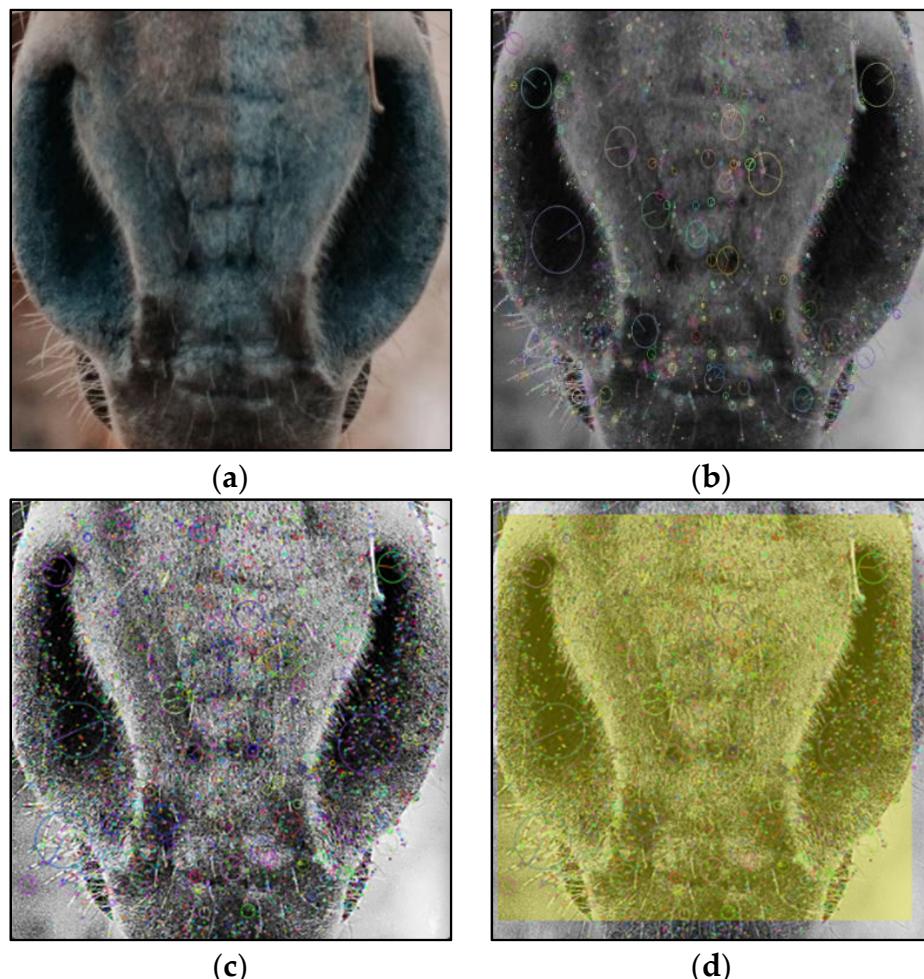
**Table 3.** Confusion Matrix of Yolo Object Detection Model for test set.

n_Samples = 242		Predicted				
		Face	Neg	Nose	Nose_	BGFP
True	Face	123	-	-	-	-
	Neg	-	8	-	-	3
	Nose	-	-	99	-	-
	Nos_	-	-	-	8	-
	BGFP	-	1	-	-	242

### 3.3. SIFT—Scale-Invariant Feature Transform

To recognize objects that are complex in appearance, such as the muzzle patterns of equine/horses, descriptive and unique features are detected and extracted. This is the purpose of the SIFT detector in the proposed system. SIFT deals with the rotation and scale of the subject from which features are to be extracted using a scale space, thus being dependent on the orientation and scale of the image, which means that if the animal is far away from the camera and the face of the horse is slightly rotated, and the system can capture a non-blurry image of the muzzle, SIFT will be able to extract features at this scale. Although the Yolo object detector is trained in such a way that if there is any occlusion, the detector will not detect the nose part, and this was considered to maintain the maximum features of the nose, if any occlusion is present, the SIFT detector can still extract features from the visible subject area on the image. SIFT also ignores the luminance of the subject. It does so by detecting interest points based on blob detection on interest/keypoints or features. The keypoint for SIFT is an area on the image that is rich in its content such as brightness and color variation, and a certain uniqueness is required in this area for it to be a keypoint, which then can be used for matching; being rich in content should also mean that it is well defined for SIFT to generate a descriptor for this point, and its position in the image should also be well defined. All these are characteristics of the grooves, ridges, and beads present on the muzzle, which are detected by the SIFT detector. Figure 7b shows the keypoints on the muzzle of the horse, whereas Figure 7a shows the original image of the nose, which was detected by the Yolo object detector when the animal was quite close to the camera, and the camera was also able to capture many details of the nose. Note that

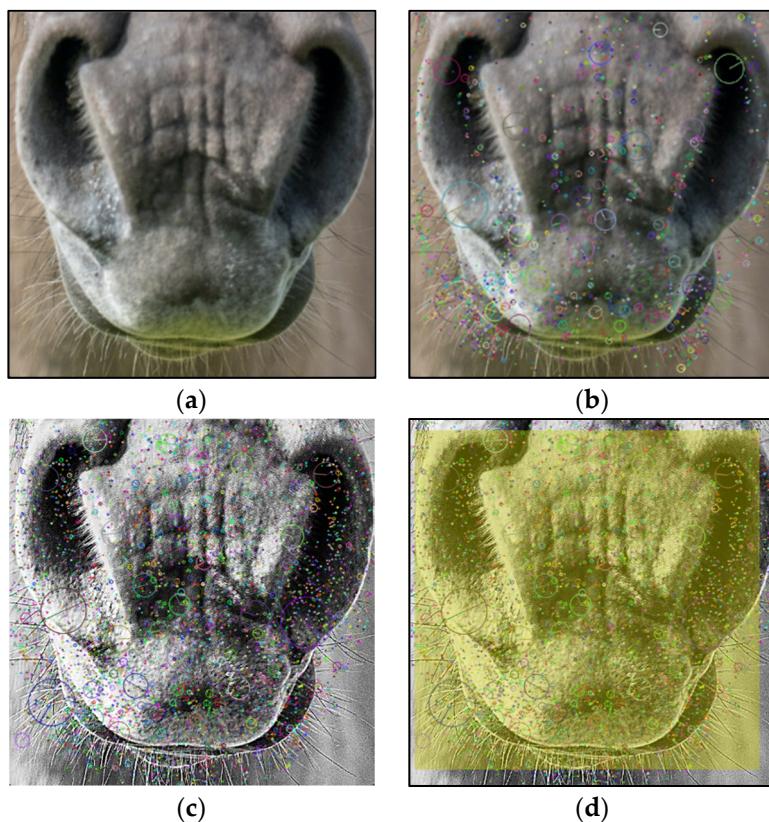
the image is converted to a gray color. Nonetheless, not all the features of the nose are detected by SIFT; hence, a Gaussian blur filter with histogram equalization is applied to the image and then the image is passed through the SIFT detector, which results in the image in Figure 7c.



**Figure 7.** Keypoint detection on muzzle of a horse: (a) originally detected by Yolo, (b) keypoint detection before processing, (c) keypoint detection after processing, and (d) region of interest is highlighted, where unique and descriptive features are detected as keypoints.

Here, we can see that the SIFT detector has detected a significantly large number of keypoints, which are manually highlighted by a rectangle in Figure 7d, which is also the region of interest. The algorithm will automatically use these keypoints along with their descriptors for matching purposes because these keypoints have unique and descriptive features that will be used for matching or identification.

Moreover, when the animal further away, or the capturing device has low quality, the following can be seen in Figure 8, where the nose of a horse was detected that was somewhat far away from the camera. We can see that the SIFT detector detected the keypoints before and after preprocessing shown in Figure 8b,c, respectively.



**Figure 8.** (a) Originally detected muzzle image of a horse far away from the camera, (b) a grayscale image with some keypoints detected by SIFT before processing, (c) unique, descriptive features detected by SIFT after processing, and (d) region of interest highlighted, where unique keypoints are detected.

#### Principle of SIFT

SIFT is a popular feature detection algorithm that is commonly used in computer vision applications. SIFT utilizes a cascade filtering approach to identify keypoints, which are distinctive and descriptive features that can be used for matching or identification purposes. The first step in keypoint detection using SIFT is to locate blobs, which are patches with local appearance, and their scale.

To locate blobs, SIFT uses a function,  $L(i, j, \sigma)$ , which convolves with a varying Gaussian,  $G(i, j, \sigma)$ , and an input image,  $I(i, j)$ , to identify blob locations over multiple scales. The varying Gaussian serves as a filter to detect blobs at different scales, while the input image is the image being analyzed for keypoints. By convolving these two elements, SIFT is able to detect blobs and their scale, which are important for keypoint detection:

$$L(i, j, \sigma) = G(i, j, \sigma) * I(i, j) \quad (1)$$

Here,  $*$  denotes the process of convolution on image  $I(i, j)$  in  $i$  and  $j$ , and  $G(i, j, \sigma)$  is the following:

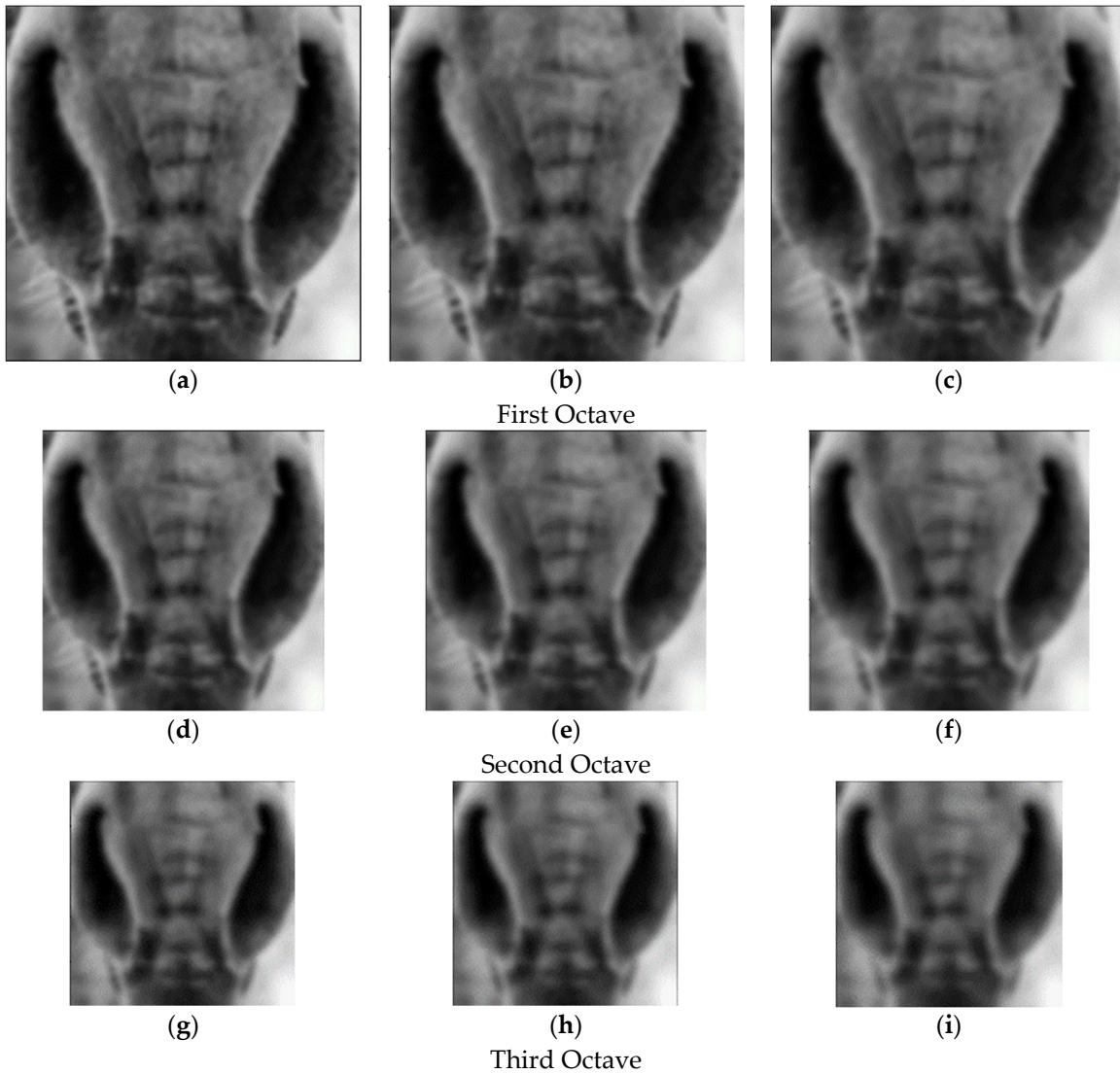
$$G(i, j, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(i^2+j^2)}{2\sigma^2}} \quad (2)$$

A scale space is created using three-octave stages while continuously blurring out the input image and dividing its size by each octave; the scale space is defined by the function in (1).

Overall, the SIFT algorithm is effective in detecting keypoints, which are distinctive and descriptive features that can be used for matching and identification purposes. Its cascade filtering approach allows for the detection of blobs and their scale, which are

important elements in keypoint detection. This makes SIFT a useful tool for identifying keypoints in images or video for a variety of computer vision applications.

This function is seen in action in Figure 9a–c are in the first octave, where blobs with larger scales are located; then, Figure 9d–f in the second octave for blobs of medium size are located, and, lastly, Figure 9g–i in the third octave for blobs of small magnification are located.

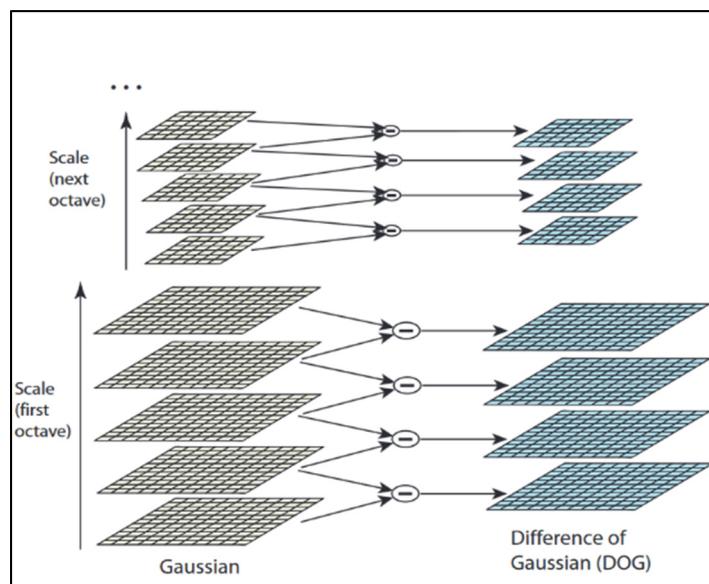


**Figure 9.** Scale-space generation using scale-invariant feature transform algorithm.

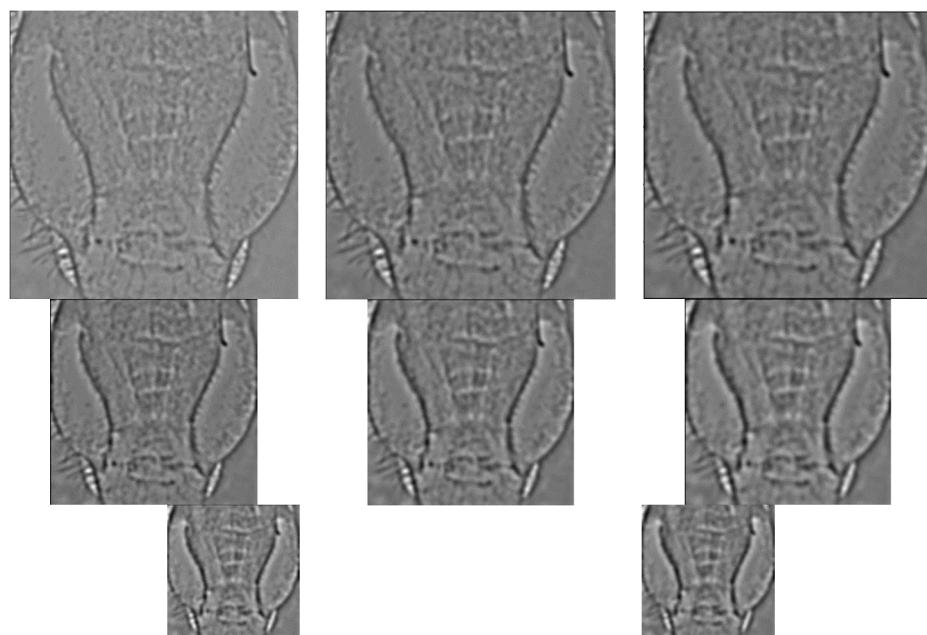
With this scale space generated, SIFT now can detect blobs, but in order to detect rich keypoints, SIFT uses the extrema of this scale space by applying convolution between difference-of-Gaussian  $D(x, y, \sigma)$  with the image. SIFT achieves this by taking the difference of nearby scales, as in Figure 10, having a difference of  $k$ , which is a constant multiple:

$$D(i, j, \sigma) = (G(i, j, k\sigma) - G(i, j, \sigma)) * I(i, j)$$

$$D(i, j, \sigma) = L(i, j, k\sigma) - L(i, j, \sigma) \quad (3)$$

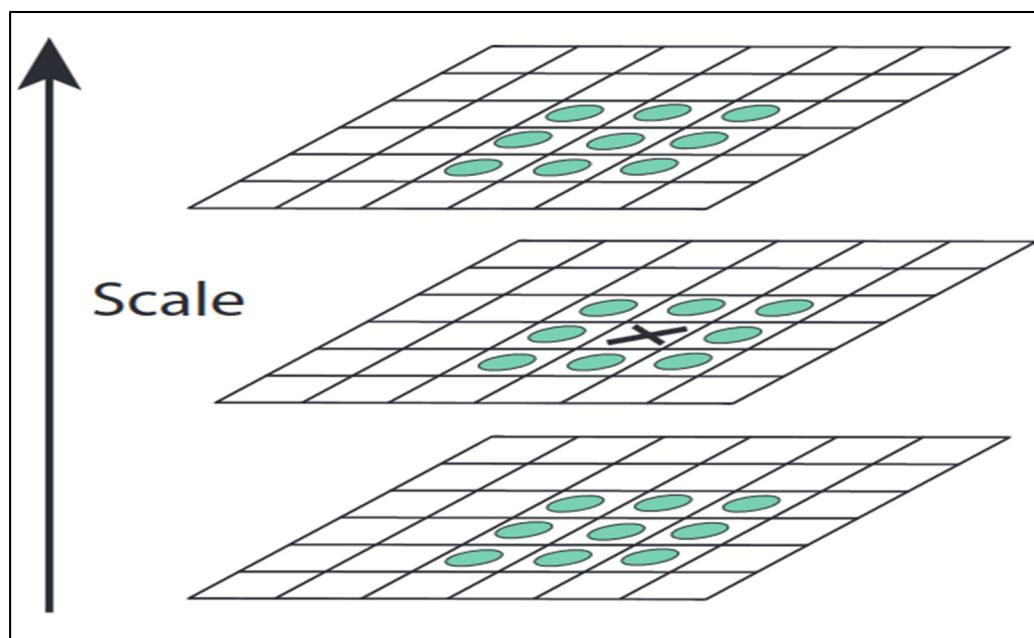


**Figure 10.** Difference of Gaussian for the detection of rich keypoints. The resulting images after passing through the difference-of-Gaussian function are shown in Figure 11, which shows the pyramid of the difference-of-Gaussian function.



**Figure 11.** Image pyramid after applying difference of Gaussian or DoG.

The blobs that either have local minima or local maxima values after passing from difference of Gaussian are considered keypoints by SIFT. SIFT finds these key-points by performing a comparison of each pixel from the image to a total of 26 neighbors, which include eight pixels present at the same scale level, nine from the scale below, and nine from the scale above; see Figure 12.



**Figure 12.** Detection of local minima and maxima.

After the detection of local minima and maxima, SIFT discards blobs that are located along the edges as the edges are not considered rich keypoints. This leaves SIFT with feature-rich keypoints localized by minima and maxima; next, SIFT assigns these points an orientation by weight of the gradient magnitude of the keypoint and then adds to a histogram with a Gaussian window with a constant scale of keypoints, and then by *argmax* of that histogram. In this way, a keypoint is created with this orientation at the location on the scale space in which it was present. The next step for SIFT is to generate a descriptor for this keypoint, which is achieved by using  $4 \times 4$  blocks from a sample of  $16 \times 16$  pixels around its location. The result of this is a descriptor vector for this keypoint. The keypoints detected by SIFT are visible in Figures 7 and 8.

We used histogram equalization from OpenCV 4.6.0 [20] to sharpen the details of the muzzle of the animal, which enabled us to detect more features. Then, this image of the muzzle of the animal is input to the SIFT detector, which returns the image with descriptors and keypoints of the muzzle of the animal.

### 3.4. FLANN

FLANN is a library used to perform approximate nearest neighbor matching in high-dimensional spaces. It is designed to be efficient and scalable, and it can be used with a variety of distance metrics and data types.

It uses two algorithms for approximate nearest neighbor matching: the randomized kd-tree algorithm and the hierarchical k-means tree algorithm. The kd-tree algorithm is a tree-based data structure that is used to efficiently search for nearest neighbors in high-dimensional spaces. It works by dividing the data into a series of nested hypercubes, and it can be used with a variety of distance metrics. The hierarchical k-means tree algorithm is a variant of the k-means clustering algorithm that is used to partition the data into a hierarchy of clusters. It is particularly well suited for large datasets with a high number of dimensions.

FLANN automatically selects the appropriate algorithm and parameters to use once the extracted features are input to it. It then returns the nearest neighbors and their distances to the query point. Two nearest neighbors are returned by FLANN in the proposed system, and the number of checks to perform and the index type to use are also defined.

Overall, FLANN is a useful tool for performing approximate nearest neighbor matching in high-dimensional spaces, and it is widely used in a variety of applications, including image and video searching, document retrieval, and machine learning.

#### 4. Results, Comparison, and Limitations

##### 4.1. Results

The proposed system, called DTL-AFI-VHA, is designed to identify animals in several stages. First, the muzzle point of the animal is extracted using a Yolo object detection model, which has demonstrated mean average precision of 99.5% for detecting the face of the animal and 99.7% mean average precision for detecting the nose or muzzle point of the animal. Next, the features of the muzzle are detected as keypoints and descriptors, which are then stored in a database. For the recognition process, FLANN is used to match the keypoints and descriptors of the animal's muzzle with those in the database. The proposed system has been tested using images from the THDD dataset and has achieved 100% accuracy in animal muzzle point feature matching. All of the animals in the THDD dataset were successfully recognized by the system.

##### 4.2. Comparisons

A comparison of the proposed system, DTL-AFI-VHA, is shown in Table 4. The proposed system has accuracy of 100% for animal identification using biometrics.

**Table 4.** Comparison of the proposed system with other studies.

Title	Technique	Accuracy %
Arabian Horse Identification and Gender Determination System Based on Feature Fusion and Gray Wolf Optimization [14]	Fusion of LBP and SURF, SVM with gray-wolf optimization	99.00
Muzzle Point Pattern Based Techniques for Individual Cattle Identification [15]	Gaussian pyramid with chi-square-based matching	93.87
Individual Beef Cattle Identification Using Muzzle Images and Deep Learning Techniques [16]	Deep learning-based multiple models	98.70
A Preliminary Investigation on Horses Recognition Using Facial Texture Features [17]	Gabor filter with LDA	95.74
Real-time recognition of cattle using animal biometrics [18]	Fisher locality preserving projection-based feature extraction	96.87
Sparse Neural Network for Horse Face Detection in a Smart Riding Club Biometric System [19]	Sparse neural network, LDA, and SVM	99.89
DTL-AFI-VHA (proposed)	A hybrid technique of automated muzzle feature extraction empowered by Yolo and SIFT, and feature matching using FLANN	100.00

##### 4.3. Limitations

One potential limitation of the proposed system for identifying and registering animals is the availability of data for all livestock animals. To train the object detection model and extract the features and descriptors of the animals' noses, a sufficient amount of data is needed. If data are not available for all of the livestock animals that the system is intended to identify, the model may not be able to accurately detect and recognize these animals. This could result in lower overall accuracy of the system and a reduced ability to identify and register all of the animals.

Another potential limitation of the system is the reliance on manual annotation of the data to label the animals' faces and noses. This process can be time-consuming and may

not be feasible for large datasets. In addition, manual annotation is subject to human error, which could also impact the accuracy of the model.

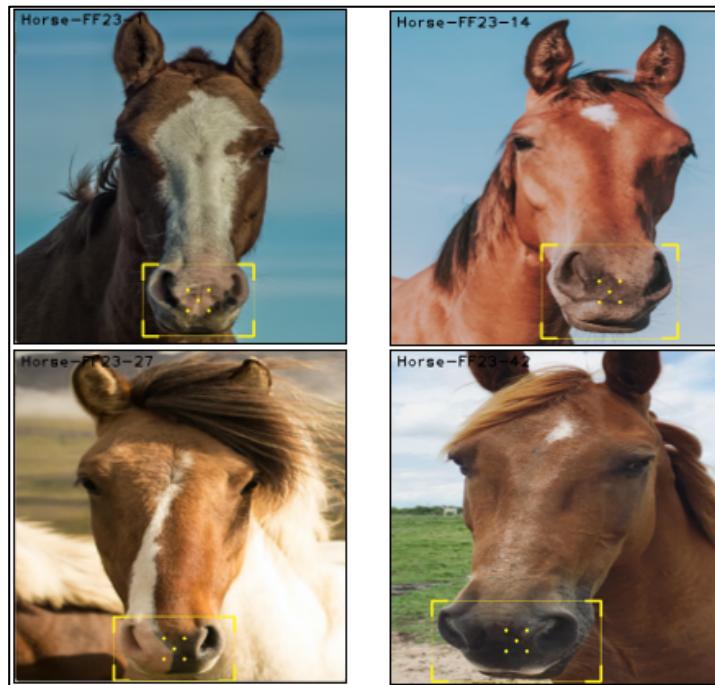
Finally, the use of approximate nearest neighbor matching with FLANN to identify the animals may also have limitations, as the method is not able to provide exact matches and is subject to a certain level of error. While the use of a threshold distance and a minimum number of good points can help to reduce the likelihood of incorrect matches, there is still the possibility of misidentification or false positives.

Overall, the proposed system for identifying and registering animals may be limited by the availability of data for all livestock animals, the reliance on the manual annotation of the data, and the potential for error in the approximate nearest neighbor matching process.

#### 4.4. Analysis of Experiments Results

In this study, we investigated the effectiveness of using Yolo v7 for face and muzzle detection, SIFT for feature extraction, and FLANN for feature matching in the identification of horses as an example of a method for identifying livestock without the use of Hot/Freeze Branding, RFID tagging and microchipping. We found that this method achieved a matching accuracy of 100%, as shown in Table 5. All the horses in our sample were correctly identified by the system. This indicates that the proposed method is a reliable and accurate way to identify horses without the need for physical contact or the use of RFID tagging or microchipping.

A few examples from horses that were successfully identified by the system during the experiment are shown in the following Figure 13.



**Figure 13.** Examples of Successfully Identified Horses.

We have analyzed the performance of the proposed model against similar research studies and found the outcome more accurate than any other methodologies previously applied. These examples demonstrate the effectiveness of the proposed identification method in correctly recognizing and matching the keypoints on the horses' noses with those in the database. The high accuracy of the system in identifying these horses highlights its potential usefulness in the livestock industry for identifying and tracking animals.

The proposed method offers several advantages for livestock identification. First, it allows for the identification of animals from a distance, which can be particularly useful in



## 5. Conclusions and Future Directions

Emerging technologies in data sciences are reshaping many cumbersome processes. AI provides a state-of-the-art solution for the automated identification of livestock. Accurate and faster biometric identification of animals has made smart farming smarter than ever before.

In conclusion, the proposed system, DTL-AFI-VHA, is a system designed to identify and register animals through the detection and recognition of their faces and noses. The system is divided into three main modules: training an object detection model, registering the animal, and identifying the animal. The object detection model is trained using deep transfer learning to accurately detect the faces and noses of animals, and the features and descriptors of the animal's nose are extracted and stored in a database. When an animal is to be identified, FLANN is used to perform approximate nearest neighbor matching of the animal's nose features with those in the database.

The purpose of this study was to develop a scalable and efficient system for identifying and registering animals using deep learning and computer vision techniques. The proposed system has demonstrated high accuracy in detecting and recognizing the faces and noses of animals, and it has the potential to be used in a variety of applications, such as livestock management, animal tracking, and conservation efforts.

Overall, the proposed system has the potential to greatly improve the efficiency and accuracy of animal identification and registration. By automating the process of detecting and recognizing animals, the system can reduce the reliance on manual methods, which can be time-consuming and prone to error. In addition, the use of deep learning and computer vision techniques allows the system to scale to large datasets and handle a wide variety of animal species. One of the main benefits of the proposed system is the ability to identify and register animals in real-time or using prerecorded videos or images. This flexibility allows the system to be used in a variety of settings and situations, such as on farms, in zoos, or in conservation areas. The achievement of optimum identification accuracy of cattle would be helpful to reduce significant losses for livestock insurers. They can better control fake claims. Moreover, cattle farmers may use the proposed model to enhance the operational efficiency of their businesses and improve animal healthcare management.

The proposed system represents a significant advancement in the field of animal identification and registration, and it has the potential to be used in a wide range of practical applications. By accurately and efficiently identifying and registering animals, the system can help to improve animal welfare, facilitate livestock management, and support conservation efforts.

Future studies should focus on how to improve the process by eliminating the limitations of data scarcity so that the relevant data for the livestock animals are available. We have to utilize an approximate nearest neighbor search with the use of a threshold distance and a minimum number of good points to reduce the likelihood of incorrect matches, but more research work is required so that this limitation can be overcome.

**Author Contributions:** M.A., S.A., and A.F. collected data from different resources. M.A., S.A., and M.A.K. performed formal analysis and simulation. M.A., G.F.I., and T.M.G. contributed to writing—original draft preparation. S.A., A.F., and M.A.K.: writing—review and editing. S.A. and M.A.K. performed supervision. M.A., A.F., G.F.I., and T.M.G. drafted pictures and tables. S.A. and M.A.K. performed revision and improved the quality of the draft. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research program is partially funded by the United Insurance Company of Pakistan Ltd. in collaboration with United Software and Technologies International (Pvt) Ltd. under Grant No. UICL-06-2022/USTI-RND-006/11/2022.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

