UTRECHT UNIVERSITY

Department of Information and Computing Science

Applied Data Science master thesis

Shape analysis of the Dutch Landrace goat based on historical imagery

First examiner:

Prof. dr. A.A. (Albert) Salah

Second examiner:

Prof. dr. M. (Mirjam) Nielen

Candidate:

Michel Doré

In cooperation with:

Landelijke Fokkersclub
Nederlandse Landgeiten

August 16, 2023

Abstract

Dutch Landrace goats, primarily kept for their cultural and heritage value, have witnessed a decline in their population and a consequent loss of distinctive aesthetics. The Dutch Landrace goat traces its ancestry to a mere four individuals in 1958, complicating efforts to retrieve its original aesthetic features. In recent years, the Landelijke Fokkersclub Nederlandse Landgeiten has attempted to address these issues. The aim of the breeders club is to preserve the Dutch Landrace goat. Obtaining more specific knowledge on what morphological characteristics were dominant is crucial for the breeding program. For this study, a dataset of historic paintings and drawings containing goats was collected and provided by the breeding club. The objective was to discover, using Procrustes analysis, whether the 2D representations derived from the dataset could aid in enhancing the breeding program, using machine learning techniques and morphological analysis. The results indicate that, while Procrustes analysis provided useful insights into the morphological variations among Dutch Landrace goats, the reliability of the outcomes was impacted by variations in artistic representations and the quality of input images. Machine learning techniques were effective in extracting goats from images, yet were limited by the quality and realism of the data. These findings underscore the potential of integrating machine learning and geometric morphometrics in historical morphology research, while also highlighting the challenges associated with data noise and reproducibility. Further research in this domain is encouraged, as the possibilities to implement new machine leaning techniques become more accessible.

Contents

1. Introduction

Unlike other goat species, the Dutch landrace goat is not primarily used for milk nor wool. The Dutch Landrace goat is considered an underdeveloped species mainly due to its low economic value and mix with other goat races. According to data from a survey at the species level, the current population can be traced back to just four Dutch landrace goats in the Netherlands in 1958 ("Food and Agriculture Organization of the United Nations," n.d.). Consequently, the current situation of the Dutch landrace goat population raises questions related to the genetic diversity and aesthetic preservation within the breeding program. This increases the difficulty of retrieving the original aesthetical features of the Dutch landrace goat, as the mixing of different goat breeds has caused these unique features to become less distinct.

(a) A current Dutch Landrace goat

(b) A goat from the dataset

Figure 1.1: Difference between current goat and dataset representation

Raising awareness and tackling these issues has been a major concern for the National Breeders Club for Dutch landrace goats (Landelijke Fokkersclub Nederlandse Landgeiten), an institute that promotes the breeding of the landrace goat and functions as a social and information point amongst breeders. The aim of the breeders club is to preserve the Dutch landrace goat. Therefore, obtaining more specific knowledge on what morphological characteristics were dominant in the original Dutch landrace goat before the turn of the 20th century is of vital importance to the breeding program. The quest for preserving the authentic Dutch Landrace goat form brings us to the research question: 'Do the 2D representations using Procrustes analysis accurately represent the head shape of the Dutch Landrace goat in comparison to the consensus shape?' Addressing this question would enable the breeders club to implement these characteristics in the current population, ensuring a more diverse and healthy breeding process that more closely aligns to the Dutch Landrace goat from the past. This research has been performed jointly with Milan van Beek. Together worked on the processing pipeline of this research. Milan's research focused on coat patterns whereas this study focused on head shape.

The significance of the research lies in its potential to improve the breeding goals of the Dutch Landrace goat breeders club, which will enhance its distinctive features. Building upon existing methodologies, this research will implement state-of-the-art machine learning algorithms, specifically a fine-tuned pretrained YoloV8 model and a model known as Segment Anything. The combined application of these models can accelerate future research in recognizing other species in both flora and fauna research, extending the implications of this study beyond the Dutch Landrace goat.

Of course, there are several limitations in this study. Firstly, the reliability of the research outcomes may be influenced by the quality of the data. The data consists mostly of paintings and drawings, of which the quality and real-world representation greatly varies. Secondly, since our understanding of the original Dutch Landrace goat's anatomy is founded on past record keeping, bias may exist in the imagery which has been chosen by the Landrace Goat breeders club. Thirdly, the machine learning algorithms employed for the detection of goats in images are sophisticated technologies; however, they may have inherent limitations and occasional inaccuracies.

This research starts by reviewing related work, to understand existing research on goat morphology, followed by a detailed explanation of the methodology used in the analysis of the goat's head shape. After the description of the methods, the results section presents the findings of the processing pipeline and morphological analysis. These findings are then discussed in a broader context in the discussion section, taking limitations as mentioned above, implications, and impact of our study into consideration. Lastly, in the conclusion section the study's findings are summarized.

2. Related Work

In the context of goat determination and detection of breeding characteristics, a few studies have been conducted which are relevant to this research. In this section the findings of these papers, and their relevance to this study will be discussed.

In a study by Getahun (2021), directly relevant to this research, 600 goats were examined. They were randomly sampled by the researchers and used for phenotypic characterization. The most prevalent coat color pattern in the sampled populations was plain and patchy with three color patterns (light red, white with red, and white) being the most frequently observed coat color type. According to Getahun (2021) "sex of animal had significant effect on all of the body measurements, except ear length, chest depth, and rump length and width". This study's findings are particularly relevant to our research as they provide insights into the phenotypic characteristics and breeding attributes of goats.

Maśko et al. (2022) has further broadened our understanding of morphological studies by comparing the shape and dorsal profiles of donkeys, ponies, and horses. Their research, primarily aimed at improving donkey welfare, utilizes an analytical method originally designed for horses. Their goal is to address posture and pain issues in donkeys, and in doing so, they emphasize the need for accurate and unbiased geometric definitions of donkey size and shape. This is seen as a significant advancement in the automated analysis of of pain and posture recognition. While their focus on animal welfare doesn't directly tie into our research, the methods used by Maśko et al. (2022) are highly relevant to our morphological study of the Dutch Landrace goat. Particularly, their application of the Procrustes analysis and Mahalanobis distance for analyzing donkey head profiles overlaps with our methodology, emphasizing its suitability for our study.

Resuming, the research conducted by Maśko et al. (2022) will aid in the field of morphology. Since their approach to calculating the head profile is one of the main pillars in our approach.

A recent study on equine pain assessment using facial images (Pessanha et al., 2022) provides valuable insights for our research. The study developed a new automatic system to predict pain in horses, using grimace scales. The process first determined how the horse's head was positioned, then pinpointed facial landmarks for categorization. Unlike our research, Pessanha et al. (2022) centered on assessing pain. They used the horse grimace scale and built a collection of horse images that were annotated by expert veterinarians. Their findings, while not directly related to goat morphology, provide an interesting application of machine learning techniques in animal welfare, which could potentially be adapted for similar applications in goat species in the future.

Furthermore, Yi et al. (2023) provides valuable insights to future research, since the majority of the data used consists of paintings. Their paper presents a novel approach to assessing the aesthetic quality of artistic images using a large-scale dataset (Boldbrush Artistic Image Dataset, BAID) and a new model (Style-specific Art Assessment Network, SAAN). In their research, Yi et al. (2023), focus on the compensation for artistic representation in artwork. This is relevant for our study.

While these studies may seem diverse in their focus areas, from the specific attributes of goats to automated pain assessment in horses, and even to the aesthetic quality of artwork, they all share a common thread of utilizing image analysis and machine learning methodologies. This demonstrates the wide-ranging applicability of such methods across different fields and contexts. Particularly, the emphasis on morphology and phenotypic characterization, as seen in the studies by Getahun (2021) and Maśko et al. (2022), is directly related to our research objectives. Furthermore, the machine learning techniques used for pain assessment in horses, as described by Pessanha et al. (2022), and the computational image analysis approach used by Yi et al. (2023) for artwork, could potentially inform future directions of our research and provide novel methodologies for the analysis of goat morphology. Although the context of the equine study might seem distant from our objective, the shared use of image analysis and machine learning methodologies bridges the gap. It shows the versatile nature of these techniques, with potential applications extending from pain detection in horses to morphological studies in goats.

In conclusion, the studies mentioned in this section contribute insightfully to this research, the field of animal and goat determination, and breeding characteristics. The findings from the mentioned papers align with the goals in our research.

3. Method

3.1 Data Preparation

In this section we will discuss the data preparation of this research. This is a vital part of this study since the dataset contained a wide variety of images. The dataset contains around 1500 images of mostly paintings and drawings dating from 15th century until the beginning of the 20th century. The dataset was collected and provided by the breeders club of the Dutch Landrace goat. The images were partly dated; some of the images included the year of creation of the painting in their filename. Since this incomplete information is not sufficient for our analysis, we left the year of creation out of the study. From a resolution standpoint, the quality of the images vary widely. A relatively large standard deviation for the width and the height compared to minimum and maximum resolutions for each axle can be seen in Figure 3.1. As for the distribution of the resolution space, this is visible in Figure 3.2.

	Width	Height
Minimum	89	77
Maximum	5092	4366
Standard deviation	754	657

Figure 3.1: Resolution information (in pixels) before filtering on image quality.

3.2 Inclusion criteria & data filtering

To move closer to analyzing the goats in the images for morphological characteristics, initial filtering is required. In consultation with goat breeding experts from the National Breeders Club of Dutch landrace goats, we established a set of requirements which should hold for each of the images in the dataset. Firstly, the images need to be realistic, which is a critical yet subjective part of the image filtering. Realistic, in this study, is defined as goats which must be represented with anatomical proportions relative to goats currently kept as livestock. The goats therefore must be a close representation of reality. Secondly, all the goats must be standing and fully visible. This is of importance for the analysis of the total shape of the goat.

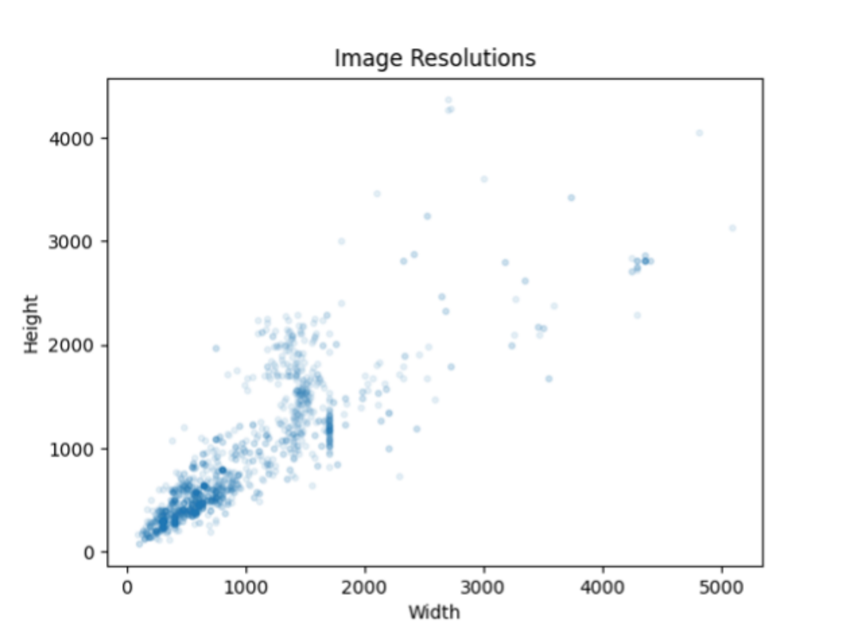

Figure 3.2: The resolution distribution of images before filtering on image quality.

Although the standing position is not necessarily critical for the analysis of the head shape, we kept this criterion due to practical constraints. Given the time limitations and various challenges in re-filtering the image dataset, it was more feasible to retain the standing posture requirement across all images, ensuring a level of consistency in our sample selection. Subsequently, every goat is similar in position making their shape comparable for the Procrustes analysis (Gower, 1975). In addition to the standing requirement, the goats must also be pictured from a side view, improving on the point made for the previous requirement. With these selection and filtering criteria in place, we moved to the next stage of our filtering process. This involved focusing more narrowly on a specific segment of the goat population: bucks.

To further enhance the reliability of the study, our selection criteria were strictly limited to bucks. This decision was largely influenced by two key factors. First, the consensus shape, which refers to the average shape that can be deduced from our Procrustes analysis, must be representative of the general goat population studied. As male goats, or bucks, often exhibit more pronounced morphological characteristics, they serve as better subjects for our study, ensuring the representability of the consensus shape. The second factor stems from the historical context of the artwork analyzed in this study. As our sources primarily include paintings, it's worth noting that these artworks overwhelmingly feature bucks. This can be attributed to their larger size and distinctive physical features, such as prominent horns, which have made them a common subject in artistic representations over centuries.

Connecting these two considerations, we found a distinctive focus on male goats, or bucks, to be a key parameter in our study. These specific selection criteria aren't merely a result of convenience, but are carefully selected based on artistic, and biological contexts.

Moving forward with the actual implementation of the labeling criteria, we developed a custom Python script. This tool enabled the goat breeders to label each image according to our specified criteria. This script was built using the tkinter library. For use in production, this script was converted to a Streamlit app which is a convenient way of running custom Python interfaces on the web. This conversion was necessary for the labelling to be completed regardless of operating system. Both scripts are available in the GitHub repository of this research, which will be mentioned in the appendix.

	Width	Height
Minimum	150	120
Maximum	4400	2999
Standard deviation	666	576

Figure 3.3: Resolution information (in pixels) after filtering on image quality.

Having established an efficient labeling method using custom Python scripts, we faced the task of applying this method to our initial diverse dataset, filtering it to get a suitable, representative sample for further analysis.

The initial dataset presented challenges due to its diversity, both in terms of image quality and the variety of depictions of goats. However, our rigorous filtering process ensured that only suitable images were selected for the study. The requirement for the goats to be realistic, standing, seen from a side view, and fully grown bucks resulted in a final sample of 82 images, which will be subjected to further analysis. This robust methodology minimizes potential sources of bias and inconsistency, ensuring that the subsequent morphological analysis is based on a carefully selected sample. This process emphasizes the importance of thorough data preparation in achiev-

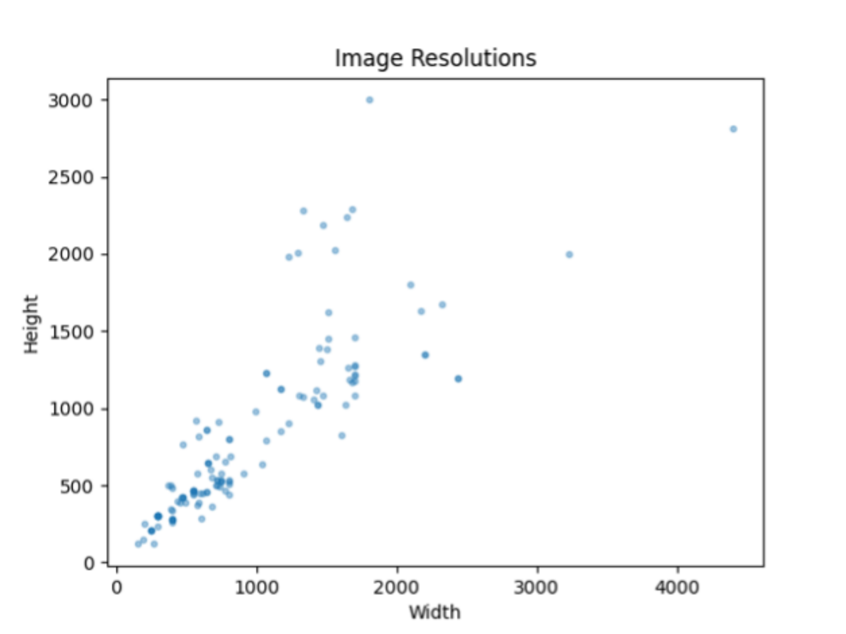

Figure 3.4: The resolution distribution of images after filtering on image quality.

ing reliable and valid results.

3.3 Processing pipeline

3.3.1 Object detection

The following step in our methodology involves the utilization of the YoloV8 model (Redmon et al., 2016). This model is an advanced deep learning algorithm known for its object detection capabilities. In this study, it is used to detect and localize goats within the images, a critical process in our morphometric analysis. Fine-tuning this model to suit our specific research needs was a meticulous task, crucial for ensuring accurate and reliable detection. In the process of fine-tuning, we started with the pre-trained YoloV8 model, which has been trained on a large dataset consisting of various real-world objects (Redmon et al., 2016). We then continued to fine-tune the model on our dataset of goat images. This allowed the model to learn the specific features of these animals, improving its ability to detect them within the images. The training images used for fine-tuning were a subset of our dataset, chosen for their clear and representative depictions. These selected images were then tagged using a bounding box at the locations where goats are present.

3.3.2 Segmentation

Once the YoloV8 model was fine-tuned to detect goats in the images, we then used the output from the YoloV8 model, specifically the bounding box information, as input to the Segment Anything model (Kirillov et al., 2023). The Segment Anything model by Kirillov et al. (2023) is a deep learning model that has been trained to segment objects from images, as can be seen at the masks and Segment Anything step in Figure 3.5 . By providing the bounding box information from the YoloV8 model, we could direct the Segment Anything model to focus on the area of the image where the goat was detected. This allowed the Segment Anything model to efficiently segment the goat from the rest of the image, leaving only the goat mask available for further analysis.

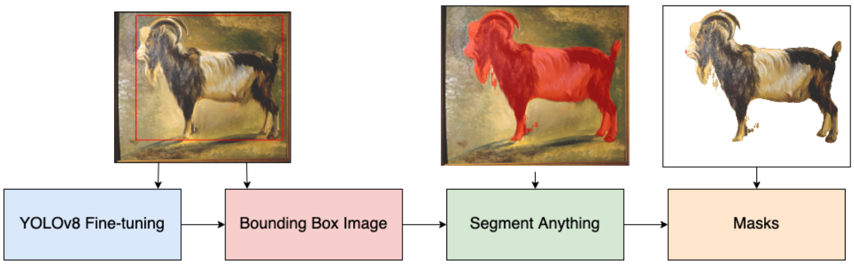

Figure 3.5: Overview of the processing pipeline.

3.4 Landmark selection

Now that we have a dataset of clear images of bucks in similar positions, we can take a closer look at their head shapes. We do this by adding landmarks, which are positions on a shape that are predefined for the research object, in our case the Dutch Landrace goat. The quality of the landmarks is a crucial aspect of conducting reliable research in geometric morphometrics. According to Watanabe (2018) it is important to ensure that the landmark data obtained from the goat masks "accurately reflect the shape variation of objects and structures under study".

To build further upon the previous claim, according to Wärmländer et al. (2018) landmark selection should be based on specific research hypotheses and justifications. Their claim emphasizes the importance of selecting landmarks based on the specific goals and questions of the study rather than general typology. Furthermore, research papers should include justifications of landmark choices along with landmark definitions (Wärmländer et al., 2018).

In our quest for the useful placement of landmarks we focused on the goat's head. Since we are looking for head shape features, we chose landmarks that identify the head shape of the Dutch Landrace goat. Specifically, landmarks were placed on both the horn bases and the tips as well as on one ear, including its bases and tip. Since only goats from a side view were used, the decision was made to only landmark the clearly visible ear. These landmarks were selected due to their consistent and clearly definable points on all goat masks.

This decision to chose the described landmarks complies with the view of Wärmländer et al. (2018), where landmarks should be selected considering specific research hypotheses and requirements. In our case, the aim is to investigate the reliability of the 2D Procrustes representations of the Dutch Landrace goat's head shape relative to the consensus shape. Therefore, each landmark has been selected after consulting specialists in this field. Which has caused each landmarks to contribute meaningfully to answering this research question and provide an accurate depiction of the distinct features of this unique goat breed.

By anchoring our landmark choices in the specific context of our research question, we enhance the robustness and reproducibility of our results. Not only does this enable us to make significant contributions to the breeding goals of the Dutch Landrace goat breeders club but it hopefully sets the stage for applying this approach to other species beyond goats.

3.5 Digitizing landmarks

The process of digitizing landmarks and converting them into vectors suitable for Procrustes analysis involved the use of tpsUtil and tpsDig2, software in geometric morphometrics for preparing and collecting landmark data.[1] Initially, all the images were added as a reference in the .tps file format using tpsUtil, a standard file format for storing coordinate data.[2] Sub-

[1] FJ, ROHLF. tpsDig2, version 2.1. http://life.bio.sunysb.edu/morph. 2006. Department of Ecology and Evolution, State University of New York at Stony Brook.

[2] Rohlf, FJ. tpsUtil, version 1.38. http://life.bio.sunysb.edu/morph/index.html. 2006. Department of Ecology and Evolution, State University of New York at Stony Brook.

sequently, tpsDig2 was utilized to place landmarks on these images. This software allows for accurate and reproducible placement of landmarks on each image based on the designated points of interest outlined in the landmark selection process.

The resulting .tps file consists of a set of coordinates for each image, corresponding to the position of the selected landmarks. This digitized data, denoting the morphology of the Dutch Landrace goat, was then ready to be incorporated into the Procrustes analysis.

3.6 Procrustes analysis

Procrustes analysis is a statistical shape analysis method used to analyze and compare geometric shapes by aligning them to a common coordinate system. The core idea behind the Procrustes analysis is to minimize the sum of squared differences between corresponding landmarks of two or more shapes. In this study, we analyzed 46 shapes. To do this, Procrustes analysis involves three main steps:

1. Translation: All shapes are centered around the origin by subtracting the centroid (mean) of their landmarks from every landmark.
2. Scaling: Shapes are scaled to have a unit size, this is achieved by dividing landmarks by the centroid size.
3. Rotation: One shape is kept fixed, while other shapes are rotated to minimize the sum of squared differences between corresponding landmarks.

For the Procrustes analysis in this study, we used MorphoJ software, a tool specifically designed for the geometric morphometric analysis of 2D and 3D shape data (Klingenberg, 2011). The software handles the alignment of landmarks by performing the Procrustes superimposition, which seeks to minimize the differences in location, scale, and orientation of the landmarks from each goat head. This superimposition provides a consensus shape and a set of Procrustes coordinates for each image.

It should be noted that there are several variants of Procrustes analysis. In this study, we specifically employed the Full Procrustes Analysis. As MorphoJ implements a full Procrustes superimposition, thereby standardizing for position, orientation, and scale to extract shape data.

Both the landmark digitization steps and the process of executing the Procrustes analysis are performed similarly to the methodology of Maśko et al. (2022).

3.7 Distance measure

Mahalanobis distance and Euclidean distance are both common distance metrics used in morphometric analysis. In this section the choice of using Mahalanobis distance over Euclidean distance will be justified.
Mahalanobis distance is a measure of distance between a point and a distribution that considers the correlations of the dataset and is scale invariant. This contrasts with the Euclidean distance, which is derived from Pythagoras' theorem and is not scale invariant. In the context of geometric morphometrics, distance measurements are often used to compare shapes. The choice of distance measure can influence results of a morphometric analysis. This choice between Euclidean and Mahalanobis distances depends on the characteristics of the dataset and the research question. (De Maesschalck et al., 2000)

In relation to this study, the Procrustes analysis generates a distinct vector for each goat by transforming the information space, as explained in section 3.6. These vectors are compared with each other for further analysis. For this purpose, the Mahalanobis distance is used which considers the distribution of landmarks to compute the distance from each point to the center of the distribution. Unlike the Euclidean distance, it does not solely measure straight-line distance. Instead, it incorporates standard deviations into the distance calculation, considering the variances within the data. For instance, a 1 cm difference in the horn length is not equal to 1 cm difference in belly size, reflecting the different scales and variances in different measurements. Therefore, the Mahalanobis distance will be used for the comparison, as it provides a more insightful and nuanced measurement of the differences between the vectors representing each goat head feature.

4. Results

4.1 Overview of the results

The findings of our study provide an exploration of the morphological characteristics of the Dutch Landrace goat as depicted in historical artwork, achieved through a modern integration of object detection, segmentation and morphological analysis.

Out of the initial dataset of roughly 1500 images, only 46 images passed our rigorous selection criteria and had sufficient and reliable information for accurate identification of specific landmarks. These images constituted our final dataset for subsequent analysis. Each image in this dataset represents a standing, fully-grown buck, depicted from a side view and meeting the realistic representation criteria as outlined in this study. To enhance the understanding of this final dataset, an overview of the resolution space is provided in Figure 4.3.

4.2 Inter-rater reliability

Assessing inter-rater reliability is essential for the validation of our data, especially since our selection and labeling processes are inherently subjective to noise to a certain extent. In order to quantify the reliability of the landmark placement, we utilized a measure known as Cohen's kappa. Cohen's kappa is a statistic that is used to measure the agreement between two raters for categorical items, and it adjusts for the probability of agreement occurring by chance (Cohen, 1960).

In this study, the inter-rater reliability assessment was performed by the same person twice, the researcher in this case, to test for noise in the landmarking. This process involved independently marking landmarks on a set of 10 images twice. A custom Python script was then used to calculate Cohen's kappa.

In order to normalize the landmarks used for calculating the Cohen's kappa a reference distance is required. There is no set reference distance for the analysis of a goat's head shape. According to Pessanha et al. (2022) for "automatically located landmarks within 10% of the inter-ocular distance to the ground truth location are considered to be accurate". Therefore we propose our own reference distance, which is defined by taking the center of the visible eye (because of the side view) to the right horn right attachment. This can be seen in Figure 4.2 This reference distance is necessary in order for the landmarks to be normalized across each of the images in the inter-rater reliability assessment. For our study, landmarks within 10% of this reference distance to the initial location were assumed as accurate.

Figure 4.1: Example of the proposed reference distance

The resulting value was 0.17, which suggests a slight agreement beyond chance according to Landis and Koch (1977) guidelines. While a higher level of agreement would be ideal, it's worth noting that there are various factors that may have influenced the level of agreement, such as the inherent subjectivity in the landmarking process and the variability in the artistic representation of the goats. A visualization is added in Figure 5, where an image with distinct landmarks denoted by red dots, representing significant points for the Procrustes analysis. The blue circle represents the error range for that specific landmark.

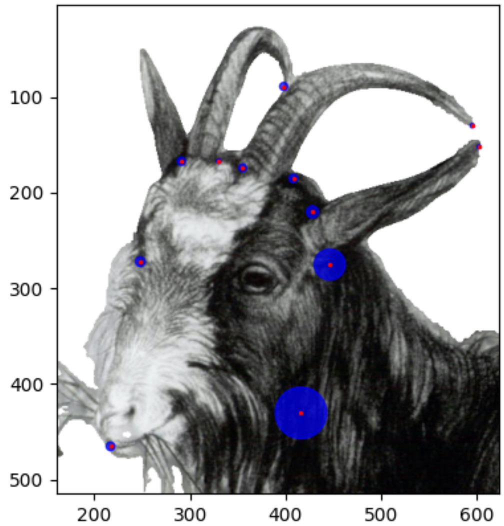

Figure 4.2: Result of the inter-rater reliability assessment.

4.3 Procrustes analysis

The Procrustes analysis was instrumental in understanding the variability in morphological characteristics of Dutch Landrace goats across our dataset of 46 images. Upon visual inspection of the Principal Component Analysis (PCA) plot (Figure 4.3), it was observed that the images were relatively scattered along the PCA1 and PCA2 axes.

PCA1 explains approximately 73.7% of the variance in our dataset, indicating a significant contribution to the overall shape variation. The variation along the PCA2 was less pronounced, contributing 25.5% to the total shape variation. These two components combined represent approximately 99.2% of the total variance in our dataset.

4.4 Mahalanobis distance & shape comparison

To assess the dissimilarity between each individual shape and the consensus shape, we computed Mahalanobis distances. This measurement, which considers the correlation of the dataset and the scale of the variables, provides a more accurate representation of the distance or dissimilarity between the shapes, beyond for example Euclidean distance.

The Mahalanobis distances plot, as represented in Figure 4.3, indicates how each individual shape deviates from the consensus shape within our dataset. The two ellipses represent two standard deviations away from the consensus shape.

Most of the images lie within the two-standard-deviation ellipse, suggesting they are close to the average depiction. However, some images fall outside this range, indicating substantial shape deviations from the consensus shape. These outliers may represent unique artistic interpretations or styles, highlighting the subjectivity and diversity in the portrayal of Dutch Landrace goats across different artists and periods.

These findings underscore the inherent variability and subjectivity in artistic representation, and caution against generalizing the morphological characteristics of Dutch Landrace goats based solely on their depiction in historical artwork.

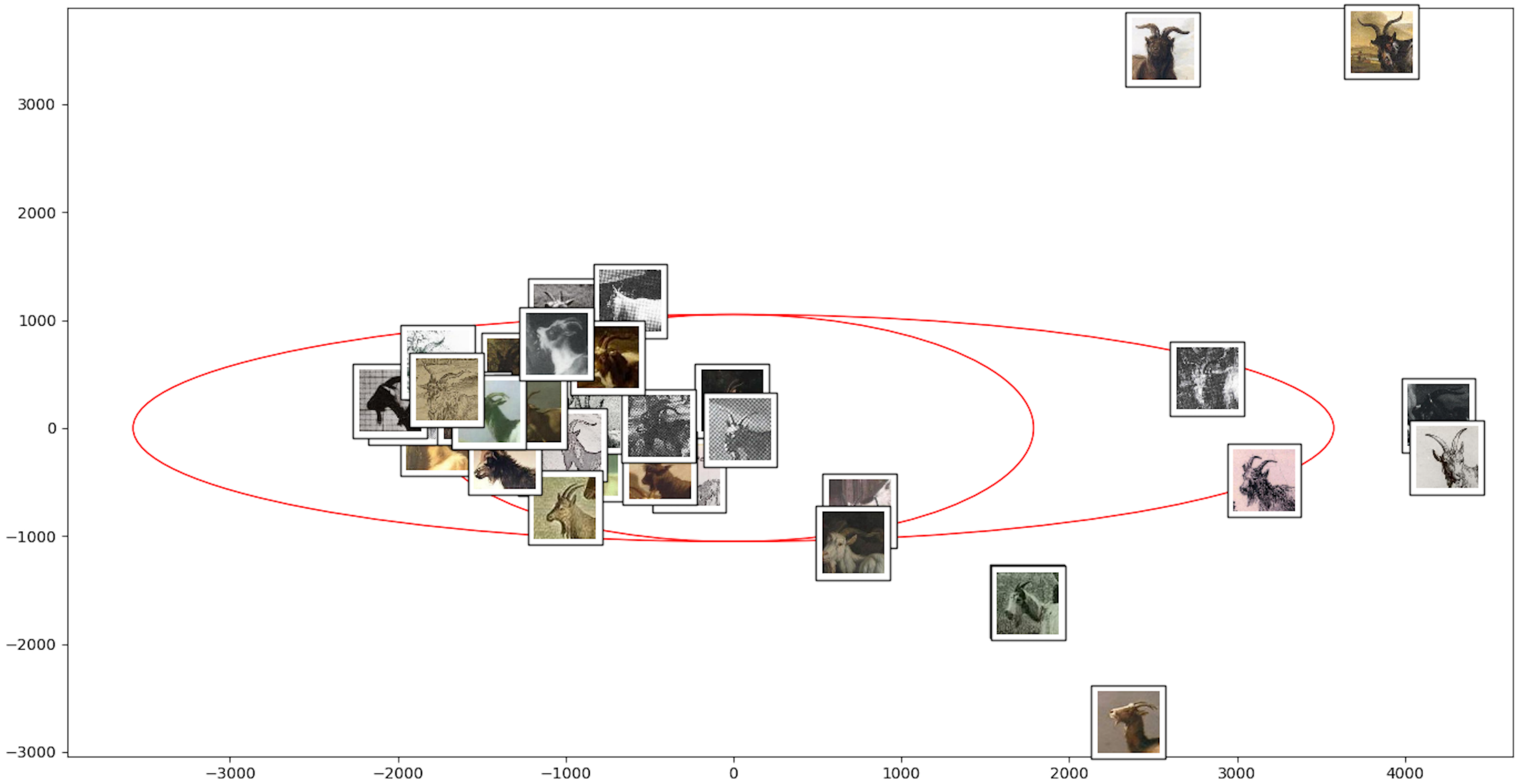

Figure 4.3: Overview of the morphological representation of goat heads using Procrustes analysis

5. Discussion

This study investigated the accuracy of 2D representations using Procrustes analysis in representing the head shape of the Dutch Landrace goat, when compared to the consensus shape. It employed advanced machine learning models and geometric morphometric methods to study the morphological characteristics of the Dutch Landrace goat from historical paintings and drawings. The results obtained from this analysis could provide important insights for the breeding programs of the Dutch Landrace goat breeders club. However, several key points need to be highlighted when interpreting the results.

Variation in artistic representations was one of the significant factors that posed a challenge for this analysis. The historical paintings and drawings used as sources of data showed considerable variations in style, angle, and realism, making the consistent identification of anatomical landmarks complicated. These variations introduced a level of noise into the dataset, which had an impact on the stability of the Procrustes analysis, as shown in the inter-rater reliability analysis. While Procrustes analysis is a powerful tool for analyzing shape and size variation and co-variation, its reliability is dependent upon accurately identifying anatomical landmarks. Our findings indicate that the variability inherent in artistic depictions, due to the subjective nature of art, make this challenging and call for further exploration of techniques to manage such noise for a more reliable analysis.

Additionally, while the YoloV8 and Segment Anything machine learning models played a critical role in extracting goat masks from images, the results of these models were heavily dependent on the quality and clarity of the input images. Less realistic or unclear depictions posed significant challenges for accurate segmentation and landmarking. This suggests a limitation of these models and points to the need for further optimization or the development of other techniques to improve the accuracy of image extraction from less clear or less realistic images. Moreover, the instability of Procrustes analysis observed in this study, as highlighted in Figure 4.2, may be attributed to this limitation. Therefore, it emphasizes the importance of exploring more robust methods to reduce noise and enhance reproducibility.

Regarding the Procrustes analysis was found to be less stable than expected. This analysis method was selected due to its ability to isolate shape information from size, position, and orientation, thus providing a clearer view of shape variations among the Dutch Landrace goats. However, the instability observed in our analysis may have resulted from the difficulties in consistently identifying anatomical landmarks. This calls for the need for more reliable methods to reduce data noise and potentially enhance the stability of such analysis. Nonetheless, Procrustes analysis still provided useful insights and remains a promising approach for future studies, particularly if these limitations can be addressed. Especially when combined with the calculation and plotting of the Mahalanobis distance.

Given these considerations, the findings of this study must be interpreted with some caution. On the other hand, this study has demonstrated the potential of combining machine learning and geometric morphometrics to investigate morphological characteristics in a historical context. It also pointed out the challenges to be addressed in refining these methodologies for more reliable and reproducible results. The outcomes of this study can be used to inform future research in this domain and potentially contribute to the improvement of breeding goals for the Dutch Landrace goat breeders club.

In conclusion, this research has contributed to a better understanding of the Dutch Landrace goat's morphological features as represented in historical artwork. It has also underscored the challenges and considerations associated with the use of machine learning and Procrustes analysis in this context. The lessons learned from this study will help pave the way for future research to further refine these techniques and enhance our understanding of historical morphological features, contributing to the preservation and restoration efforts for underdeveloped species like the Dutch Landrace goat.

6. Conclusion

The coming together of historical artwork with computational analysis, as presented in this study, underscores a new approach towards understanding and preserving unique traits of understudied species like the Dutch Landrace goat. As evidenced in the study, the 2D representations of the goat's head shape using Procrustes analysis proved to be a somewhat promising path of research, with the potential to highlight the morphological characteristics of the Dutch Landrace goat.

However, as is inherent with introductory studies, further research is needed. Conservation efforts and breeding programs for the Dutch Landrace goat would benefit greatly from the continuation of this line of research. The identification and implementation of distinct morphological characteristics could significantly enhance the genetic diversity and aesthetic appeal of this breed and probably other breeds. Yet, these conclusions must be tempered with an acknowledgment of the challenges encountered in the course of the research and the need for future studies to validate the findings.

Moreover, the dataset, composed primarily of historical paintings and drawings, raises some unique challenges. The reliability and realism of these images can be subjective, as they are the artists' interpretations of the goats. However, the employment of machine learning models in this study, specifically YoloV8 and Segment Anything models, assisted in identifying and segmenting the goats from these images, and this approach shows promise for further research in identifying morphological features in other fauna.

In conclusion, this study signifies a critical step towards using computational methods to dive into historical data and uncover essential morphological characteristics of the Dutch Landrace goat. Further research in this area is encouraged, and the methodological innovations presented here could offer substantial contributions to the preservation of the Dutch Landrace goat.

6.1 Acknowledgement

We would like to express our heartfelt gratitude to our supervisors, Albert Salah and Mirjam Nielen, for their guidance and support throughout the course of this research. And to Milan van Beek, working together was an extremely educative and positive experience. Your expertise and dedication have helped greatly in the completion of this study.

Additionally, our thanks extend to Albert van de Streek and Piet Enserink, for their contribution to this project and taking their free time to spend on this research. Their deep understanding of the Dutch Landrace goat and commitment to their conservation greatly aided our research and made this study possible.

We would also like to thank Mieke Maarse and Martie Doré, who have greatly adviced in structuring and writing correct language.

Finally, we thank all who contributed to this project in one way or another. Your collective effort has been crucial to this achievement.

Bibliography

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*(1), 1–18. https://doi.org/10.1016/s0169-7439(99)00047-7

Food and Agriculture Organization of the United Nations. (n.d.). https://www.fao.org/

Getahun, Z., Ahmed. (2021). Morphometric characterization of indigenous goats in East Gojjam zone, Amhara region, Ethiopia. *Journal of Natural Sciences Research*, *7*(2-2020). https://doi.org/10.7176/jnsr/12-17-03

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*(1), 33–51. https://doi.org/10.1007/bf02291478

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2304.02643

Klingenberg, C. P. (2011). Morphoj: An integrated software package for geometric morphometrics. *Molecular ecology resources*, *11*(2), 353–357. https://doi.org/10.1111/j.1755-0998.2010.02924.x

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Maśko, M., Wierzbicka, M., Zdrojkowski, L., Jasiński, T., Sikorska, U., Pawliński, B., & Domino, M. (2022). Comparison of Donkey, Pony, and Horse Dorsal Profiles and Head Shapes Using Geometric Morphometrics. *Animals*, *12*(7), 931. https://doi.org/10.3390/ani12070931

Pessanha, F., Salah, A. A., Van Loon, T., & Veltkamp, R. C. (2022). Facial Image-Based Automatic Assessment of Equine Pain. *IEEE Transactions on Affective Computing*, 1. https://doi.org/10.1109/taffc.2022.3177639

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Wärmländer, S. K., Garvin, H. M., Guyomarc'h, P., Petaros, A., & Sholts, S. B. (2018). Landmark Typology in Applied Morphometrics Studies: What's the Point? *Anatomical Record-advances in Integrative Anatomy and Evolutionary Biology*, *302*(7), 1144–1153. https://doi.org/10.1002/ar.24005

Watanabe, A. (2018). How many landmarks are enough to characterize shape and size variation? *PLOS ONE*, *13*(6), e0198341. https://doi.org/10.1371/journal.pone.0198341

Yi, R., Tian, H., Gu, Z., Lai, Y.-K., & Rosin, P. L. (2023). Towards artistic image aesthetics assessment: A large-scale dataset and a new method. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22388–22397.