Article

Automated Processing and Phenotype Extraction of Ovine Medical Images Using a Combined Generative Adversarial Network and Computer Vision Pipeline

James Francis Robson [1]*, Scott Denholm [2] and Mike Coffey [2]

[1] Scotland's Rural College (SRUC), Peter Wilson Building, Kings Buildings, West Mains Road, Edinburgh EH9 3JG, UK.   james.robson@sruc.ac.uk
[2] Scotland's Rural College (SRUC), Peter Wilson Building, Kings Buildings, West Mains Road, Edinburgh EH9 3JG, UK.
* Correspondence

Abstract: The speed and accuracy of phenotype detection from medical images are some of the most important qualities needed for any informed and timely response such as early detection of cancer or detection of desirable phenotypes for animal breeding. To improve both these qualities, the world is leveraging artificial intelligence and machine learning against this challenge. Most recently, deep learning has successfully been applied to the medical field to improve detection accuracies and speed for conditions including cancer and COVID-19 [1,2]. In this study, we applied deep neural networks, in the form of a generative adversarial network (GAN), to perform image-to-image processing steps needed for ovine phenotype analysis from CT scans of sheep [3,4]. Key phenotypes such as gigot geometry and tissue distribution were determined using a computer vision (CV) pipeline. The results of the image processing using a trained GAN are strikingly similar (a similarity index of 98%) when used on unseen test images. The combined GAN-CV pipeline was able to process and determine the phenotypes at a speed of 0.11 seconds per medical image compared to approximately 30 minutes for manual processing. We hope this pipeline represents the first step towards automated phenotype extraction for ovine genetic breeding programmes.

1. Introduction

Increase in global food demand has led to livestock breeders seeking to produce breeding lines more able to match economic demand which have genetic advantages to primary traits such as growth speed and reduced feed intake. With agricultural animals providing 18% of global calories and 39% of global protein intake, they are still an essential part of global nutritional requirements [5]. One of the methods in making livestock more advantageous is to selectively breed them for commercial traits such as growth rate [6], milk quality [7] , weather [8] and disease resistance [9]. Recent improvements in genomic technologies such as detection of single nucleotide polymorphisms (SNPs) and whole genome sequencing   [10] have allowed unparalleled insight into the driving factors which guide animal phenotypes [11] and successful genomic breeding selection has been able to identify traits which are not only desirably economically, such as improved livestock social behaviour and carcass composition [9], but also identify novel cosmetic or welfare indicators such as predicting horn phenotypes in Merino sheep [12]. As the number and biological complexity of known phenotypes is increasing, there is a call to innovate new ways to detect phenotypes faster and more accurately [13] in addition to detecting and preserving those of potential future relevance [14].

Non-invasive imaging techniques, such as computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound, can provide detailed data from which phenotypes can then be extracted [15,16] and used in breeding programmes. One major benefit of using these non-invasive imaging techniques is that internal phenotypic data, such as muscle and fat distribution [17], organ size and limb morphology, can then be incorporated swifter into genetic breeding programmes for live breeding animals [18]. Out of the commonly used non-invasive imaging techniques, CT scanning provides the highest resolution (1-2mm). One hurdle which can impact extraction of useful phenotypic information is the processing and analysis of these images which can be time consuming and therefore costly, especially if there is a need to re-analyse historic databases to measure newly emerging phenotypes.

Machine learning and artificial intelligence have been successfully implemented to increase phenotype detection speed and accuracy within many different medical areas including brain cancer detection, covid status in lungs and classification of organ deformities [1,2,19]. Recently the same technology has been applied to areas of agricultural science such as detection of bovine tuberculosis status based upon milk spectral data (Denholm et al., 2020; Brand et al., 2021). Briefly put, these networks work by passing data such as images, or segments thereof, through a series of layers containing artificial neurones which determine the likelihood of visually similar animals such as pigs, sheep or horses on a scale of 0 (absent) to 1 (present). The type of network commonly used to perform this image to binary diagnostic is a convolutional neural network where, as the layer depth increases, many data-points (such as pixels) are condensed into fewer data-points (likelihood of e.g. pig, sheep or horse presence). The subject field of artificial intelligence, machine learning and deep learning using neural networks is extremely broad, and this research article only aims to provide a broad overview in order to demonstrate its application in agriculture and to not discuss these in-depth, although many excellent reviews exist for further reading [22–25].

To perform image to image translations a similar type of neural network is required, although rather than condensing pixel information into a few datapoints, the shape of the layers more closely resembles that of an hourglass laying on its side (Figure 1a). This hourglass shape allows the network to perform general purpose image-to-image translation and even increase resolution of blurry input images [26]. By pairing this image-to-image transforming network with a second convolutional neural network (Figure 1a), the discriminator, which compares and scores the images produced by the image transforming network and tries to discriminate between fake and ground truth results, a self-training system can be produced. These two-component image translational networks are termed generative adversarial networks (GANs) and have traditionally been used for a variety of image translational tasks including sketch-to-photo, smile-to-frown, and nonbearded-to-bearded [26–28]. More recently, GANs have been applied to medical images to remove noise from low-dose CT, generate tissue structure from blood vessel networks, correct motion artefacts, produce CT images from MR images, and synthesise new image data [29–31].

By combining GANs with another machine learning technique, computer vision (CV), any images generated by the GAN can then be analysed to extract data of interest in a fully automated way (Figure 1b). Computer vison is a research field which aims to extract understanding or context from images and can use both traditional mathematical regression techniques as well as deep learning classification networks [32,33]. Application of CV can range from simple inspection of food quality and ripeness by counting the number of pixels within images of fruit and vegetables which fall within certain colour hue ranges [34,35] up to more complex tasks such as identifying road signage or pedestrians to guide automated driving systems [36].
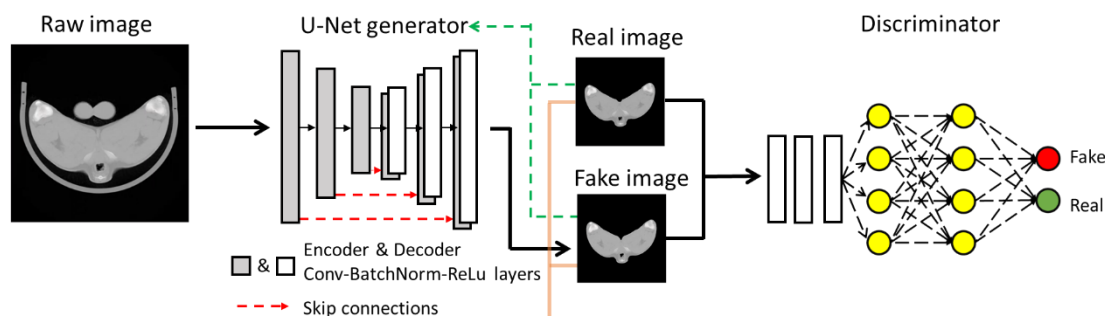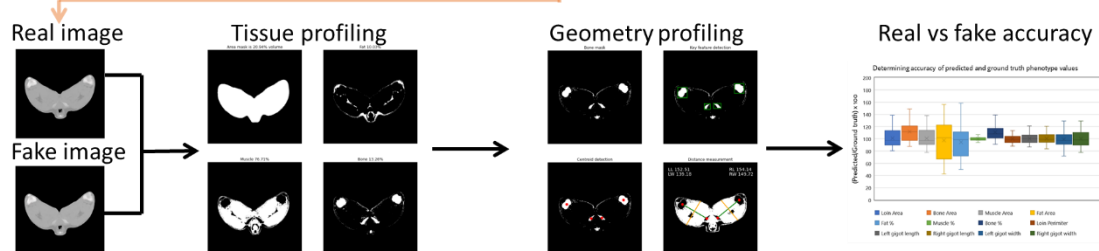
Figure 1. Combined GAN-CV pipeline for phenotype extraction. Neural networks can be trained to perform image to image translations such as in a) where a raw ovine CT scan is passed through a generator network, a series of convolution, batch normalisation and ReLu activation function layers, to produce a "fake" image. Skip connections apply regions from the encoded to the encoded images and improve object border definition. By reducing differences between the real and the fake images (green dashes) the autoencoder also learns to better produce fake images independantly. A second neural network, the discriminator, then determines if an image is considered real or fake. By pairing the two neural networks to work against each other, an adversarial component emerges, where the generator tries to produces images to fool the discriminator and the discriminator tries to determine if these images are real or not.

We use both smart techniques (GANs and CV, Figure 1a & b respectively) to aid processing and analysis of agricultural medical images of sheep. This research aims to first implement a GAN to perform ovine CT processing steps involving global information manipulation such as object and organ removal since within the image are multiple objects (scanning cradle and padding) and organs (testes) of varying size, morphology and orientation. Then, with the processed image containing only key features, attempt to extract phenotypes relevant for breeding programmes using CV techniques in an automated process.

2. Materials and Methods

2.1 Ovine ischium scan collection

A single cross-sectional 2D images was taken through the top of the leg at the point of the ischium for each lamb using a Siemens Somatom Scope with a slice thickness of 10mm for a variety of breeds including Beltex, Blue Texel, Charollais, Hampshire Down, Meatlinc, Shropshire, Southdown, Suffolk and Texel as performed by Bunger et al [16]. The images at this stage are referred to as "raw" images throughout the paper as they are unprocessed. All CT images produced from the scans are exported in the "Digital Imaging and Communications in Medicine" (DICOM) format, a unified filetype for medical imaging techniques. Such DICOM images contain additional data regarding the subject, such as age, sex, and location, in addition to collection parameters such as equipment and scanning methodology used. Image dimensions used for this study were 512x512 pixels of an intensity value between 0 (black) and 2550 (white) where 0 typically represents low density matter such as air and 2550 represents extremely dense matter such as metal.

2.2 Determination of tissue pixel intensities

Pixel intensities corresponding to respective tissues of fat, muscle and bone were calculated based upon comparison with dissected tissue as explored by Bunger et al [16]. Pixel intensity windows for fat, muscle and bone were 800-1000,1000-1100 and 1100-1750 respectively. This groups previous research allowed us to incorporate set pixel intensity windows for each tissue type into the CV pipeline easily.

2.3 Manual image processing and phenotype analysis

All images had been previously labelled by manual phenotype extraction. Parts of the image superfluous for downstream phenotype determination including scanning cradle and testes were removed using STAR software routines [16] using the method described by Glasbey et al [37]. Images produced from this processing are considered as "ground truth". From the ground truth images, tissue phenotype could then be extracted by calculating tissue distribution within the experimentally determined windows. Other phenotypes such as gigot length were measured manually by measuring the distance from the centre of the ischium bone cross-section to that of the femur bone cross-section in a "click and drag" fashion. Processing the images in this fashion took approximately 30 minutes.

2.4 GAN model

GANs are two component systems which have a generator component G to generate images and a discriminator component D to determine if the image is real or fake. The generator G takes an input image to translate into an output image y and can operate in either an unconditional fashion where random noise z is supplied or in a conditional fashion where an input image x or random noise z is supplied, $G : \{x \text{ or } z\} \rightarrow y$. The discriminator D determines if the image produced is "real" or "fake" and helps train the generator G to produce images which can pass as "real". GANs thus attempt to optimize the following function [38]:

$$\min_{G} \max_{D} V(G,D) = \mathbb{E}_{x,y} [\log D(x,y)] +$$

$$\mathbb{E}_{x,z} [\log (1 - D(x, G(x,z)))] \tag{1}$$

Further improvement of the generator G can be incorporated by including a function to minimise the absolute pixel differences between "real" and "fake" images [27].

$$L_{L1}(G) = \mathbb{E}_{x,y,z}[y - G(x,z)] \tag{2}$$

Which results in the following final model:

$$G^* = \min_{G} \max_{D} V(G,D) + L_{L1}(G) \tag{3}$$

2.4.1 GAN training

The GAN network trained in this study is an implementation of AUTOMAP [39] and Pix2Pix [27] which has been optimised for use with paired image datasets [3] available online at (https://github.com/shizuo-kaji/PairedImageTranslation). This particular GAN was chosen for this study as it was designed from the ground up to process paired sets of images, such as those commonly found in the medical field where an image can be altered to produce a "before" and "after" whilst maintaining the same subject ID and type e.g. sheep-sheep, human-human in a conditional synthesis process. This is in contrast to other popular GANs, such as CycleGAN and DCGAN, which perform unconditional synthesis by capturing key style concepts, from large batches of example images to translate images between two highly different abstract style concepts such as horse-to-zebra, photograph-to-Van Gogh or sketch-to-cat [40,41].

A dataset containing 126 raw and ground truth image pairs of mixed breed ovine CT scans taken between 2019-2020 were used for GAN training (Supplementary file 1). DI-COM pairs were first split into training (n = 101) and validation (n = 25) datasets (80% and 20% respectively). The raw and ground truth pairs of DICOM filename IDs were first given a suffix of "_0" or "_1" respectively to act as identifiers. All file extensions were then modified to ensure compatibility with the DICOM processing libraries used in this study. The script used to train the GAN, along with the full list of GAN settings used for this study is available within Supplementary file 2. Key settings for training the GAN were as follows: random translation = 0, epochs = 100, weight for L1 reconstruction loss = 0, weight for L2 reconstruction loss = 10.0, weight for softmax focal reconstruction loss = 1.0, weight for total variation = 1e-3. Following training, both the L1 (absolute pixel difference) and L2 (mean squared error) were approaching stable values (Supplementary Figures 1).

2.5.1 Image processing using trained GAN on unseen data

32 raw CT scans (Supplementary file 3) taken between 2018-2019 and belonging to the breed Charollais were passed through the trained GAN model to produce "predicted" images that were given a suffix of "_2" to clearly differentiate between the raw and ground truth counterparts (Supplementary file 2).

2.6 CT scan similarity comparison

2.6.1 CT scan histogram comparison

Alternative image manipulation techniques, such as removing pixels above or below certain intensities, were not suitable for processing the CT scans in the DICOM format as the pixel intensities of image objects needing to be removed overlapped with that of the subject's tissue. Furthermore, pixels in certain areas could not be removed since subject orientation was not constant. Due to the large irregular pixel area changes needed to process the images a deep neural network that can perform image to image translations was deemed to be of potential use. This can be visualised by comparing the pixel intensity histograms of both the Raw and Ground truth images below in figure 2 (generated as part of the computer vision pipeline in Supplementary file 4).
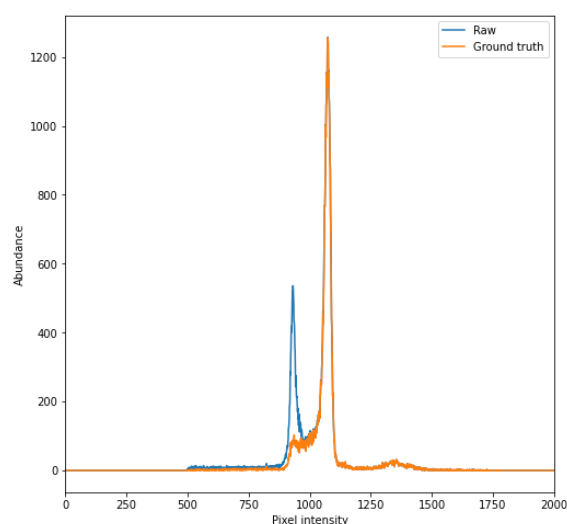
Figure 2. A representative pixel intensity histogram of raw and ground truth image shows large variance. By comparing the raw and ground truth pixel intensity histograms it can be visualised that they a) share certain areas of similarity (as seen at the peak between 1000 and 1250) but also b) contain regions which have different non-zero abundances (within the peak between 750 and 1000). As there are no regions where pixel intensity is either present or not present, images cannot be processed by simply flattening

pixel intensities which lie between certain values. This type of non-linear transformation is a task in which neural networks perform well.

2.6.2 Calculation of image similarity

Mean squared error (MSE) and structural similarity index (SSI) metrics were used to compare the raw and ground truth images with the resulting predicted images. Mean squared error is a full pixel-wise reference metric with values closer to zero being better; it is the sum of the accumulative mean squared difference across each pixel location between a pair of images. this technique, however, is extremely sensitive and seemingly large amounts of MSE can be accumulated by very minor shifts in the image, as perceived by the human eye, such as slight rotations or horizontal and vertical translations [42]. A newer, more wholistic and subtle approach which avoids the extreme position sensitivity of MSE is calculating the SSI, which analyses local similarities in structure, luminance and contrast to closer mimic how the human eye perceives similar images [43]. Both MSE and SSI was calculated for each pairwise comparison of image classes (raw, ground truth or predicted in this study) using the SciKit Image python image processing library as documented in Supplementary file 4 [44].

2.7 Phenotype measurement using computer vision

Automated phenotype extraction from ground truth and predicted (processed) images was performed using a pipeline which incorporated known pixel intensity value thresholding for each component of the carcass, based upon manual dissection, for each tissue type in combination. Geometric phenotypes were computed predominantly using the area, contour and perimeter functions within the CV library SciKit Image [44]. In addition, a set of bespoke functions were also written to detect probable tissue pixel intensities windows of fat, muscle and bone if no known set values were available, or if the images being analysed were from different sources. All steps of phenotype extraction using computer vision are documented in Supplementary file 4.

2.7.1 Tissue distribution

The area of all tissues within the ground truth and predicted images were calculated using the SciKit Image contour function for later use in determining percentage tissue composition. Tissue masks for each image were applied by first setting pixel intensity values (Fat muscle bone) outside the respective tissue windows to zero and then setting values within the window to max (2550). Fat, muscle, and bone % of each image was determined by comparing the number of pixels that fell within each of the respective tissue masks to that of the area of all tissue. By visualising each of the tissue masks independently, muscle and fat distribution could be observed in addition to locations of key physical features such as bones for further geometric phenotype analysis.

2.7.2 Skeleton geometry

One key phenotype used for estimation of muscularity is the ratio of the length and width of the gigot muscle. These dimensions are typically measured by hand from the CT scan image but by using CV models, we can extract this information automatically from the bone tissue mask image by implementing SciKit Image area and crofton perimeter functions [45]. Since small pieces of grit and sand may appear in the bone mask, due to high density as detected by x-rays, only bone mask objects over 200 pixels in both area and perimeter were referenced. Then, to avoid including spinal bone tissue, the four largest objects in the most +Y direction are assumed to be the features of interest and are placed into pairs according to their position along the X axis. The distance in pixels is then calculated between each pair of bones to determine gigot length. A line perpendicular to

that between the bone pairs is then used to find the furthest non-zero positions within the muscle tissue mask and thus determine gigot width.

2.8 Computing hardware and software

The training of machine learning models can be an intensive computational task which typically requires powerful graphics processing units (GPUs). As such, all computation was performed on a NVIDIA DGX Station workgroup server [46]. The DGX workstation provided supercomputing performance with one out of a total of four TESLA V100 GPU being used for computations underpinned by an Ubuntu operating system. All code was run within a Compute Unified Device Architecture (CUDA) 10.1 docker container which allows parallelisation of general-purpose processing to be applied to the powerful GPUs. Within this container, the open source learning framework Chainer was used to accelerate creation of the neural networks [47]. The GAN trained in this study is an implementation of AUTOMAP [39] and Pix2Pix [27] which has been optimised for use with paired image datasets [3] available online at (https://github.com/shizuo-kaji/PairedImageTranslation). Predicted images produced by the GAN were then processed using a bespoke python script run within a Jupyter notebook (Supplementary file 4). The notebook contains code within cells which can either a) run individual steps and generate intermediary output figures (slower) or b) calculate metrics and compare images without visualising any medical images (faster).

3. Results

The trained model was able to transform the raw images with a high degree of accuracy and perform the large image area manipulations, such as scanning cradle and testicle removal, needed to produce images similar to the manually processed ground truth images. The accuracy of these transformations was confirmed by visual inspection of predicted images and measurement of image similarity metrics including MSE and SSI. Phenotypic traits such as fat, muscle and bone tissue distribution and both gigot length and width were then automatically extracted from the predicted (transformed) images using CV techniques. All values calculated using this pipeline area are recorded in an output file (Supplementary file 5)

3.1 CT scan processing using trained GAN

Raw CT scans not previously seen by the GAN were processed using the trained model at a speed of 0.11 seconds per scan. Predicted and ground truth images and pixel intensity histograms were first compared visually to initially assess GAN suitability and ensure that they were visually similar (Figure 3). Quantitative metrics such as MSE and SSI were further determined to accurately assess the success of the GAN for processing the CT scans (Figure 4).

3.1.1 Images produced from trained model.

The trained model was able to perform the major structural alterations within the image dataset needed to transform the raw CT scans into something which, by eye, strongly resembled the ground truth images as shown below in Figure 3. Image IDs 1732, 9638 and 8353 were chosen to illustrate this transformation since, on visual inspection, as they contained the highest area of features needing to be removed (large testes and a large scanning cradle).

310
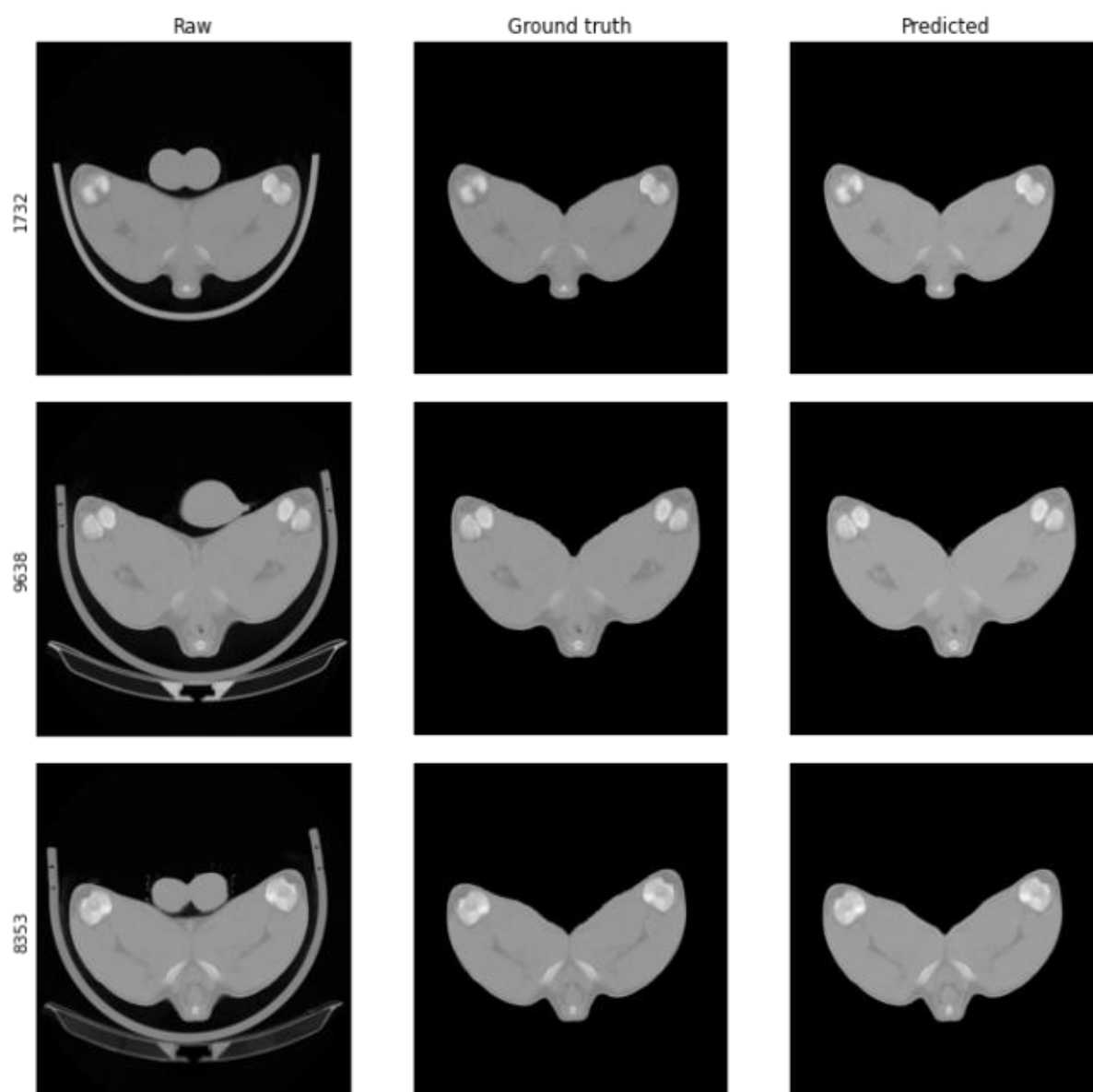
Figure 3. Representative comparison of raw, ground truth and predicted CT scan images. A trained Generative Adversarial Network (GAN) was used to process raw CT images (left column) into something resembling manually processed ground truth images (middle column). Non-quantitative visual inspection of predicted results (right column) indicate that images produced by this GAN are similar to ground truth counterparts. The GAN showed good capabilities in automatically handling the large image transformations needed to remove image objects such as testes and scanning cradle.

3.1.2 Image similarity metrics confirm a high degree of similarity.

Just as raw and ground truth image histograms were compared previously, likewise the ground truth and predicted images were compared in a similar fashion which revealed two histograms, highly similar, showing a large proportion of overlap and a high degree of similarity from visual inspection. Likeness of the raw, ground truth and predicted image sets (n= 32) were compared pairwise using MSE and SSI. Both raw vs ground truth and raw vs predicted showed the lowest image similarity values with an average MSE of (58674 ± 17766 and 58008 ± 17319) and with average SSIs of (0.49 ± 0.025 and 0.48 ± 0.024) respectively indicating a high degree of image dissimilarity. On the other hand, comparing images in the ground truth and predicted datasets showed a much lower

average MSE (1028 ± 1201) and a far higher average SSI of (0.98 ± 0.0035) indicating a far greater similarity and indicated high accuracy of the trained model in mimicking the manual processing of CT scan images.
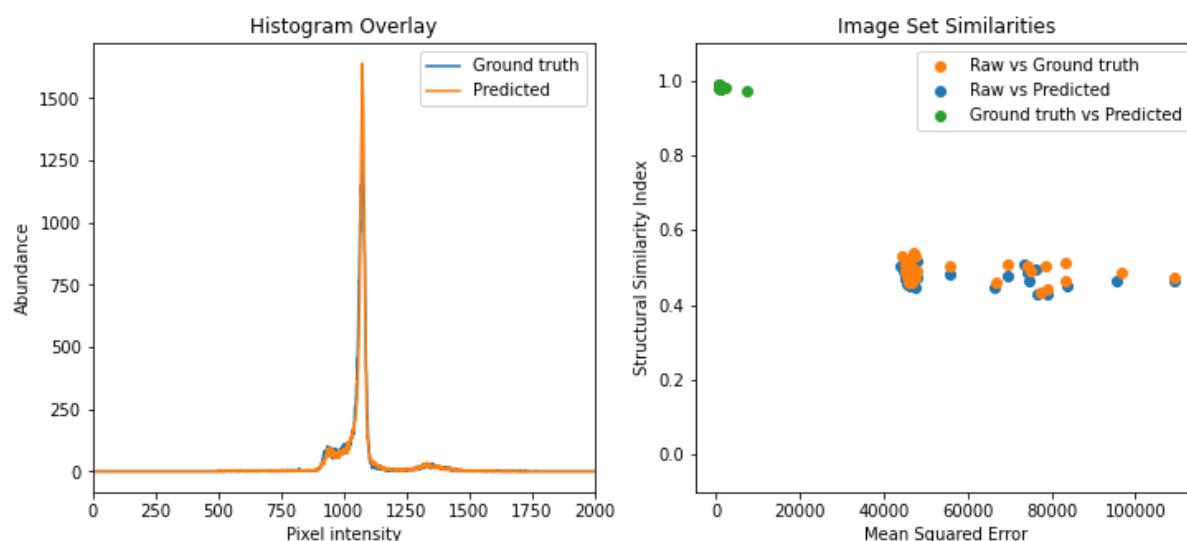

Figure 4. Quantifying a high degree of quantified similarity between ground truth and predicted images. Comparing a representative pixel intensity histogram of a ground truth and predicted image (left) showed a high degree of overlay and that peaks were present in similar areas at similar amplitudes indicating a similar distribution of pixel intensities within each image. Structural components of image groups were compared (right) using mean squared error (MSE) and structural similarity indexes (SSI) which revealed a) high average MSE (58674 ± 17766 and 58008 ± 17319, n = 32) with low average SSI (0.49 ± 0.025 and 0.48 ± 0.024, n = 32) between raw vs ground truth and raw vs predicted image groups respectively b) low average MSE (1028 ± 1201) and high average SSI (0.98 ± 0.0035) when comparing ground truth vs predicted images. These high SSI and low MSE values confirm the suitability of a trained generative adversarial network to perform highly accurate ovine CT image processing.

## 3.2 Automated phenotype extraction

The image processing library SciKit Image was successfully implemented to provide CV capabilities in the automated phenotype extraction pipeline. In this study, phenotypes of interest included fat, muscle and bone tissue abundance as well as leg geometry such as length and width (Supplementary file 4).

3.2.1 Leg tissue composition

Tissue abundance and distribution of fat, muscle and bone, within the single 2D image analysed, was calculated by counting pixels which fell within experimentally determined tissue pixel intensity windows compared to the total tissue area. Binary visualisation of these tissue value windows allowed rapid profiling of tissue distribution as seen below in Figure 5. Using this method, tissue abundances were calculated for each medical image in terms of both area and percentage composition (Figure 6). On average, the area of bone, muscle and fat across the dataset was 6488 ± 533, 44274 ± 4051 and 5712 ± 1377 mm^2. Carcass tissue composition percentage-wise for bone, muscle and fat was 11.52 ± 0.78, 78.41 ± 1.90 and 10.07 ± 2.03 %.
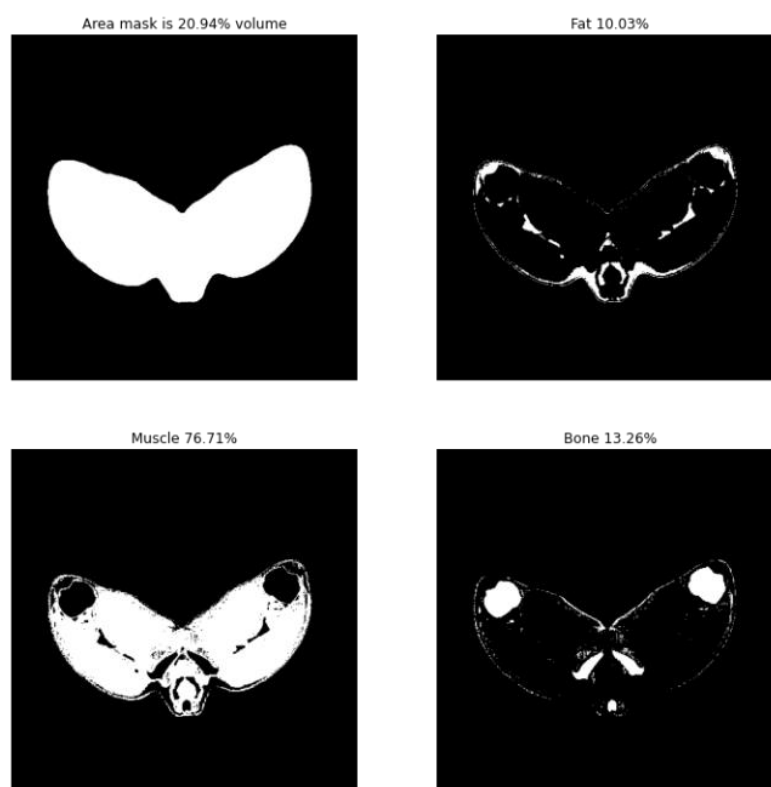
Figure 5. Representative tissue distribution of fat, muscle and bone within tissue area of predicted images. The results of the trained generative adversarial network were analysed by examining total area (top left) and by applying pixel intensity threshold windows to separately visualise fat (top right), muscle (bottom left) and bone (bottom right). The total number of pixels that fell within these pixel intensity windows determined the volume of the respective tissue types in the sample since 1 pixel = 1 mm^2.
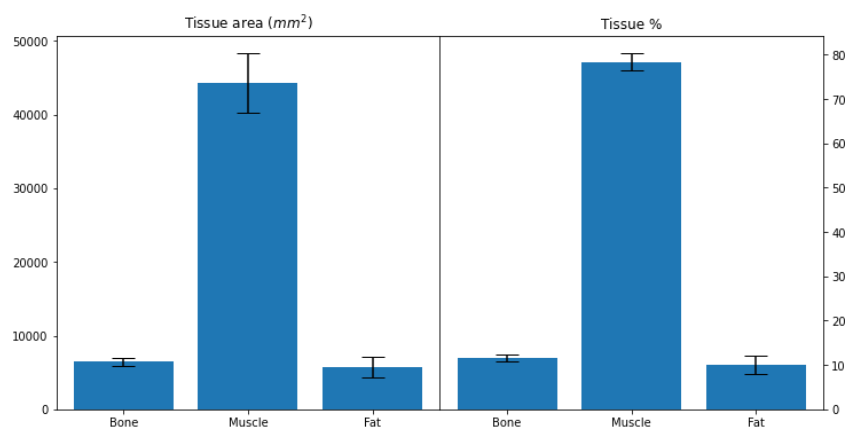
Figure 6. Area and percentage composition of tissue types within predicted ovine medical images. Using the threshold windows for each of the respective tissue types, the total area occupied was calculated for each tissue type (left) and what percentage this represented within each individual CT scan (right). On average, the area of bone, muscle and fat across the dataset was 6488 ± 533, 44274 ± 4051, 5712 ± 1377. Carcass tissue composition percentage-wise for bone, muscle and fat was 11.52 ± 0.78, 78.41 ± 1.90 and 10.07 ± 2.03 %.

3.2.2 Gigot length and width phenotype extraction. 368

By applying CV functions from the SciKit Image library such as area, perimeter and 369
location restraints to objects in the bone tissue mask, the position and centre of key fea- 370
tures was detected, and gigot length and width determined automatically as part of the 371
CV script (Supplementary file 4). This process can be visualised below in figure 7. Left 372
and right gigot lengths were 164.45 ± 8.72 mm and 166.38 ± 9.71 mm with widths being 373
137.55 ± 10.53 mm and 143.99 ± 12.42 respectively. 374

375
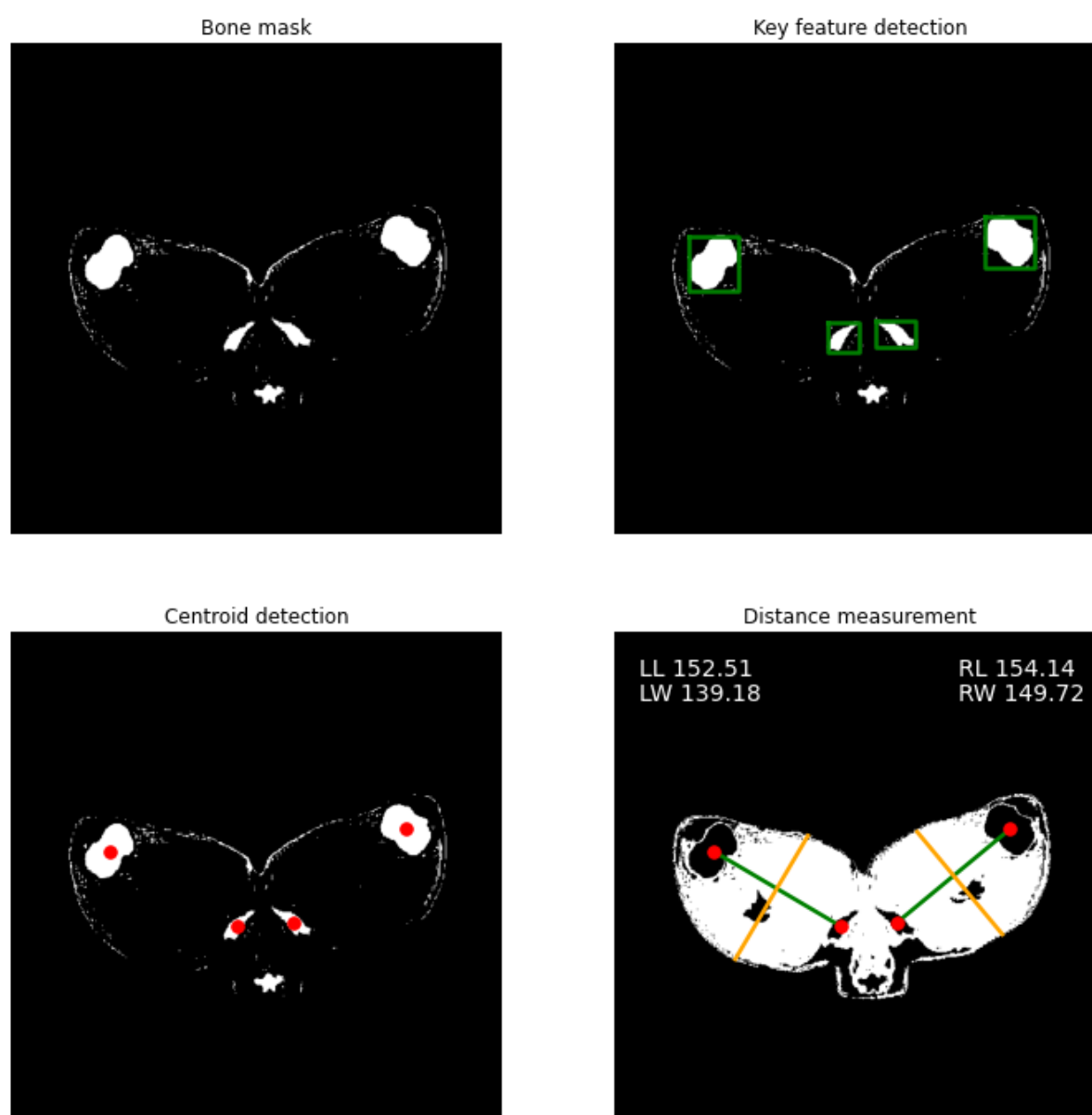
Figure 7. Automated identification of key features and determination of gigot length and width from predicted images. Using 376
the bone tissue mask (top left) from the predicted image generated by the GAN key features can be identified (top right). The centroid 377
of each object within each feature area can be calculated (bottom left) to then measure distances and determine gigot length (LL,RL 378
bottom right). By taking the perpendicular equation of the linewhich connects the two pairs of bones, the width of the gigot (LW,RW, 379
bottom right) can be calculated by discovering the first and last non-zero values of these positions within the muscle tissue mask. 380
Left and right gigot lengths on average were 164.45 ± 8.72 mm and 166.38 ± 9.71 mm with widths being 137.55 ± 10.53 mm and 143.99 381
± 12.42 respectively. 382

383

3.2.3 Phenotype extraction accuracy

Phenotypes from both predicted and ground truth datasets were extracted using the computer vision pipeline and compared to determine the suitability of predicted images for phenotype determination as seen below in Figure 8. Across all phenotypes, the average values were on average 101.44 % that of the ground truth value with a standard deviation of 12.90 % (n = 32). Muscle % was the most accurate predicted phenotype with estimated values between 93.67 and 106.65 %. On the other hand, calculating fat area was the least accurate predicted phenotype with estimated values between 42.50 and 156.18 % (Following incomplete ovine testes removal from image ID 8346, fat related phenotypes were not included in accuracy calculations as testes are calculated as fatty tissue. All other phenotypes for this image were recorded normally such as muscle area, bone area and gigot geometry).
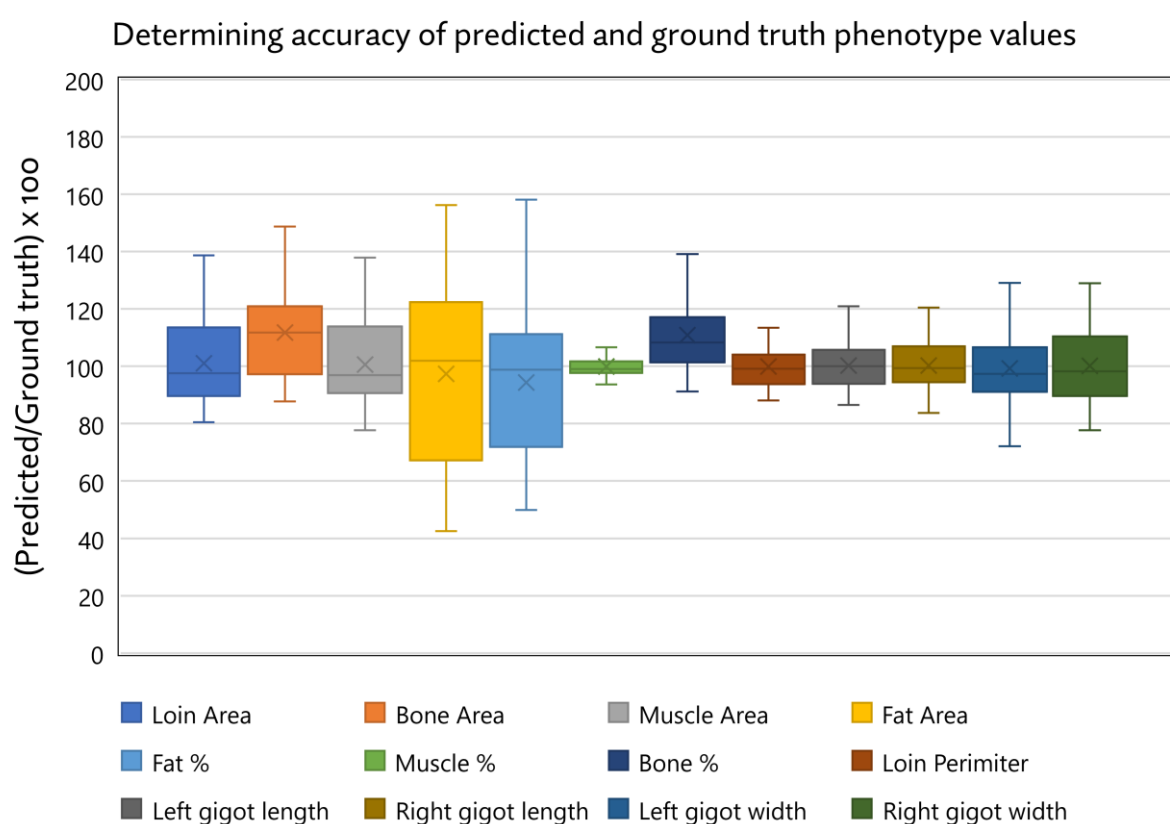
Figure 8. **Comparing values of predicted and ground truth phenotypes.** Prediction estimation accuracy was determined by comparing phenotype values generated from both predicted and ground truth datasets using the computer vision pipeline. Accross all phenotypes, predicted values for each image were on average 101.44 % that of the ground truth value with a standard deviation of 12.90 % (n = 32)

4. Discussion

Continued reduction in DNA genotyping cost over time has resulted in mainstream integration of genomic selection into genetic improvement programmes for a number of domesticated animals. The increase in availability of genotypes leads to the need to identify the correlated phenotypes, as subtle or rare as they may be [14]. One technology which shows great promise in detecting these subtle phenotypes is the use of trained neural networks and CV. The processing and extraction of key data from medical images in the past has been typically performed manually by trained and experienced professionals. However, more recently, emergence of trained artificial intelligence networks have contributed

to increased analysis throughput and accuracy of phenotype determination, such as the increased use and accuracy of neural networks for cancer and disease detection compared to the results of medical professionals [1,2,48,49]. By implementing similar techniques into the field of animal breeding, we hope to enhance the speed and accuracy of phenotype detection to streamline swift integration into genetic improvement programmes.

As part of this automated pipeline, a generative adversarial network was first trained to perform the necessary image to image translation required for automatically processing previously unseen CT scan images for subsequent phenotype extraction using CV at a speed of 0.11 seconds per image, this speed is far greater than the approximate 30 minutes required to manually process the image. The resultant images processed in this manner had an SSI of (0.98 ± 0.0035) when compared to the manually processed ground truths according to their structural similarity index and were visually indistinguishable. Automated phenotype extraction from predicted CT images was then performed by subdividing each image into the respective tissue masks to display the fat, muscle and bone volume and distribution. By using key feature detection within the bone image mask, distances between ischium and femur bone cross-sections were calculated to determine the geometric phenotype of gigot length and width. Phenotype values determined using the computer vision pipeline were on average 101.44 % that of the ground truth value with a standard deviation of 12.90 % (n = 32) indicating a high level of accuracy across the population.

One of the potential limitations of this study was the small training data set (m = 126) as development of neural networks typically uses datasets numbering in the thousands. However, this limited dataset did not cause any major issues in accuracy as ground truth and predicted images showed an SSI of (0.98 ± 0.0035) and were indistinguishable by eye. One possible reason for such high accuracy with this limited dataset was that all subjects within the CT scans were constrained to similar postures. This hypothesis was later confirmed by re-introducing artificial random movement (such as rotation or vertical/horizontal shifts) into the images used for GAN training, resulting in a higher validation loss, poorer network performance and blurry resultant images (Supplementary figures 1-9).

Unfortunately using this limited dataset resulted in one of the unseen images containing a small amount of testes tissue following processing with the GAN which was then incorrectly quantified as fat tissue. In the future, as more images are integrated into the model, we believe that the accuracy of GAN shall improve which shall directly improve the precision of the CV phenotype determination pipeline.

5. Conclusion

In summary, we believe this research represents the first case of using an automated phenotype detection pipeline on agricultural animal medical images. This was achieved by using a combined GAN-CV pipeline to analyse agricultural medical images in a fully automated fashion. By feeding a paired image dataset into a GAN, we were able to perform the various image processing steps needed to produce a predicted image, containing only the relevant tissues, with accuracies of 98% which rivalled that of manual processing and at a fraction of the cost. Phenotypes were then extracted or calculated from these predicted images by applying CV techniques as part of an automated pipeline.

We hope to immediately expand this highly accurate GAN-CV pipeline to process and extract phenotypes from other key CT scan sections such as the 8th thoracic vertebrae and 5th lumbar vertebrae positions. Further on, we hope to develop a pipeline to process a complete set of layered CT images to produce an accurate 3D model from which a multitude of phenotypes can then be extracted, such as spine length and vertebrate number, and detect phenotypes which are best explored in 3D space such as organ morphology [50,51]. By continuing this research we will further expand the automated extraction of phenotypes from agricultural medical imaging data and use the findings to guide genetic and genomic breeding programmes

5. References

1. Savage, N. Savage-2020-How-ai-is-improving-cancer-diagnost. **2020**.

2. Lin, L.; Qin, L.; Xu, Z.; Yin, Y.; Wang, X.; Kong, B.; Bai, J.; Lu, Y.; Fang, Z.; Song, Q.; et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* **2020**, *296*, E65–E71, doi:10.1148/radiol.2020200905.

3. Kaji, S.; Kida, S. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiol. Phys. Technol.* **2019**, *12*, 235–248, doi:10.1007/s12194-019-00520-y.

4. Tripathy, S.; Kannala, J.; Rahtu, E. Learning Image-to-Image Translation Using Paired and Unpaired Training Samples. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2019**, *11362 LNCS*, 51–66, doi:10.1007/978-3-030-20890-5_4.

5. FAO Shaping the future of livestock (Report No. I8384EN). *10th Glob. Forum Food Agric.* **2018**, 18–20.

6. Rexroad, C.; Vallet, J.; Matukumalli, L.K.; Reecy, J.; Bickhart, D.; Blackburn, H.; Boggess, M.; Cheng, H.; Clutter, A.; Cockett, N.; et al. Genome to phenome: Improving animal health, production, and well-being - A new USDA blueprint for animal genome research 2018-2027. *Front. Genet.* **2019**, *10*, 1–29, doi:10.3389/fgene.2019.00327.

7. Gonzalez-Recio, O.; Coffey, M.P.; Pryce, J.E. On the value of the phenotypes in the genomic era. *J. Dairy Sci.* **2014**, *97*, 7905–7915, doi:10.3168/jds.2014-8125.

8. Sánchez-Molano, E.; Kapsona, V. V.; Ilska, J.J.; Desire, S.; Conington, J.; Mucha, S.; Banos, G. Genetic analysis of novel phenotypes for farm animal resilience to weather variability. *BMC Genet.* **2019**, *20*, 1–10, doi:10.1186/s12863-019-0787-z.

9. Brito, L.F.; Oliveira, H.R.; McConn, B.R.; Schinckel, A.P.; Arrazola, A.; Marchant-Forde, J.N.; Johnson, J.S. Large-Scale Phenotyping of Livestock Welfare in Commercial Production Systems: A New Frontier in Animal Breeding. *Front. Genet.* **2020**, *11*, 1–32, doi:10.3389/fgene.2020.00793.

10. Li, X.; Yang, J.; Shen, M.; Xie, X.L.; Liu, G.J.; Xu, Y.X.; Lv, F.H.; Yang, H.; Yang, Y.L.; Liu, C. Bin; et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat. Commun.* **2020**, *11*, 1–16, doi:10.1038/s41467-020-16485-1.

11. Santos, B.F.S.; Van Der Werf, J.H.J.; Gibson, J.P.; Byrne, T.J.; Amer, P.R. Genetic and economic benefits of selection based on performance recording and genotyping in lower tiers of multi-tiered sheep breeding schemes. *Genet. Sel. Evol.* **2017**, *49*, 1–16,

doi:10.1186/s12711-016-0281-2.

12. Duijvesteijn, N.; Bolormaa, S.; Daetwyler, H.D.; Van Der Werf, J.H.J. Genomic prediction of the polled and horned phenotypes in Merino sheep. *Genet. Sel. Evol.* **2018**, *50*, 1–11, doi:10.1186/s12711-018-0398-6.

13. Seidel, A.; Krattenmacher, N.; Thaller, G. Dealing with complexity of new phenotypes in modern dairy cattle breeding. *Anim. Front.* **2020**, *10*, 23–28, doi:10.1093/af/vfaa005.

14. Leroy, G.; Besbes, B.; Boettcher, P.; Hoffmann, I.; Capitan, A.; Baumung, R. Rare phenotypes in domestic animals: Unique resources for multiple applications. *Anim. Genet.* **2016**, *47*, 141–153, doi:10.1111/age.12393.

15. Han, D.; Lehmann, K.; Krauss, G. SSO1450--a CAS1 protein from Sulfolobus solfataricus P2 with high affinity for RNA and DNA. *FEBS Lett.* **2009**, *583*, 1928–32, doi:10.1016/j.febslet.2009.04.047.

16. Bunger, L.; Macfarlane, J.M.; R., N.; Conington, J.; A., K.; Moore, K.; Glasbey, C.A.; Simm, G. Use of X-Ray Computed Tomography (CT) in UK Sheep Production and Breeding. *CT Scanning - Tech. Appl.* **2011**, 329–348, doi:10.5772/23798.

17. Lee, S.; Lohumi, S.; Lim, H.S.; Gotoh, T.; Cho, B.K.; Jung, S. Determination of intramuscular fat content in beef using magnetic resonance imaging. *J. Fac. Agric. Kyushu Univ.* **2015**, *60*, 157–162, doi:10.5109/1526313.

18. McLaren, A.; Kaseja, K.; McLean, K.A.; Boon, S.; Lambe, N.R. Genetic analyses of novel traits derived from CT scanning for implementation in terminal sire sheep breeding programmes. *Livest. Sci.* **2021**, *250*, 104555, doi:10.1016/j.livsci.2021.104555.

19. Lim, L.J.; Tison, G.H.; Delling, F.N. Artificial Intelligence in Cardiovascular Imaging. *Methodist Debakey Cardiovasc. J.* **2020**, *16*, 138–145, doi:10.14797/mdcj-16-2-138.

20. Denholm, S.J.; Brand, W.; Mitchell, A.P.; Wells, A.T.; Krzyzelewski, T.; Smith, S.L.; Wall, E.; Coffey, M.P. Predicting bovine tuberculosis status of dairy cows from mid-infrared spectral data of milk using deep learning. *J. Dairy Sci.* **2020**, *103*, 9355–9367, doi:10.3168/jds.2020-18328.

21. Brand, W.; Wells, A.T.; Smith, S.L.; Denholm, S.J.; Wall, E.; Coffey, M.P. Predicting pregnancy status from mid-infrared spectroscopy in dairy cow milk using deep learning. *J. Dairy Sci.* **2021**, doi:10.3168/jds.2020-18367.

22. Soffer, S.; Ben-Cohen, A.; Shimon, O.; Amitai, M.M.; Greenspan, H.; Klang, E. Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. *Radiology* **2019**, *290*, 590–606, doi:10.1148/radiol.2018180547.

23. Hou, X.; Gong, Y.; Liu, B.; Sun, K.; Liu, J.; Xu, B.; Duan, J.; Qiu, G. Learning based image transformation using convolutional neural networks. *IEEE Access* **2018**, *6*, 49779–49792, doi:10.1109/ACCESS.2018.2868733.

24. Bod, M. A guide to recurrent neural networks and backpropagation. *Rnn Dan Bpnn* **2001**, *2*, 1–10.

25. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine learning in agriculture: A comprehensive updated review. *Sensors* **2021**, *21*, 1–55, doi:10.3390/s21113758.

26. Parmar, N.; Vaswani, A.; Uszkoreit, J.; Kaiser, L.; Shazeer, N.; Ku, A.; Tran, D. Image transformer. *35th Int. Conf. Mach. Learn. ICML 2018* **2018**, *9*, 6453–6462.

27. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* **2017**, *2017-Janua*, 5967–5976, doi:10.1109/CVPR.2017.632.

28. Zhong, G.; Gao, W.; Liu, Y.; Yang, Y.; Wang, D.H.; Huang, K. Generative adversarial networks with decoder–encoder output noises. *Neural Networks* **2020**, *127*, 19–28, doi:10.1016/j.neunet.2020.04.005.

29. Singh, N.K.; Raza, K. Medical Image Generation using Generative Adversarial Networks. *arXiv* **2020**, 1–19.

30. Armanious, K.; Jiang, C.; Fischer, M.; Küstner, T.; Hepp, T.; Nikolaou, K.; Gatidis, S.; Yang, B. MedGAN: Medical image translation using GANs. *Comput. Med. Imaging Graph.* **2020**, *79*, 101684, doi:10.1016/J.COMPMEDIMAG.2019.101684.

31. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **2018**, *321*, 321–331, doi:10.1016/J.NEUCOM.2018.09.013.

32. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review.

Comput. Intell. Neurosci. **2018**, *2018*, doi:10.1155/2018/7068349. 550

33. Meer, P.; Mintz, D.; Rosenfeld, A.; Kim, D.Y. Robust regression methods for computer vision: A review. *Int. J. Comput. Vis.* 551
    1991, *6*, 59–70, doi:10.1007/BF00127126. 552

34. Wu, D.; Sun, D.W. Colour measurements by computer vision for food quality control - A review. *Trends Food Sci. Technol.* 553
    2013, *29*, 5–20, doi:10.1016/j.tifs.2012.08.004. 554

35. Brosnan, T.; Sun, D.W. Improving quality inspection of food products by computer vision - A review. *J. Food Eng.* **2004**, *61*, 555
    3–16, doi:10.1016/S0260-8774(03)00183-3. 556

36. Deva Koresh, J. Computer Vision Based Traffic Sign Sensing for Smart Transport. *J. Innov. Image Process.* **2019**, *1*, 11–19, 557
    doi:10.36548/jiip.2019.1.002. 558

37. Glasbey, C.A.; Young, M.J. Maximum a posteriori estimation of image boundaries by dynamic programming. *J. R. Stat. Soc.* 559
    Ser. C Appl. Stat. **2002**, *51*, 209–221, doi:10.1111/1467-9876.00264. 560

38. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative 561
    adversarial networks. *Commun. ACM* **2020**, *63*, 139–144, doi:10.1145/3422622. 562

39. Zhu, B.; Liu, J.Z.; Cauley, S.F.; Rosen, B.R.; Rosen, M.S. Image reconstruction by domain-transform manifold learning. *Nature* 563
    2018, *555*, 487–492, doi:10.1038/nature25988. 564

40. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial 565
    networks. *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.* **2016**, 1–16. 566

41. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 567
    Proc. IEEE Int. Conf. Comput. Vis. **2017**, *2017-Octob*, 2242–2251, doi:10.1109/ICCV.2017.244. 568

42. Sara, U.; Akter, M.; Uddin, M.S. Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *J.* 569
    Comput. Commun. **2019**, *07*, 8–18, doi:10.4236/jcc.2019.73002. 570

43. AGandhi, S.; C.V.Kulkarni MSE Vs SSIM. *Int. J. Sci. Eng. Res.* **2013**, *4*, 930–934. 571

44. Van Der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Gouillart, E.; Yu, T. Scikit-image: 572
    Image processing in python. *PeerJ* **2014**, *2014*, 1–18, doi:10.7717/peerj.453. 573

45. Séverine, R. Analyse d′image géométrique et morphométrique par diagrammes de forme et voisinages adaptatifs généraux. 574
    2011. 575

46. NVIDIA NVIDIA DGX Station: AI Workstation for Data Science Teams Available online: https://www.nvidia.com/en- 576
    gb/data-center/dgx-station-a100/ (accessed on May 21, 2021). 577

47. Tokui, S.; Okuta, R.; Akiba, T.; Niitani, Y.; Ogawa, T.; Saito, S.; Suzuki, S.; Uenishi, K.; Vogel, B.; Vincent, H.Y. Chainer: A 578
    deep learning framework for accelerating the research cycle. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **2019**, 579
    2002–2011, doi:10.1145/3292500.3330756. 580

48. Lassau, N.; Ammari, S.; Chouzenoux, E.; Gortais, H.; Herent, P.; Devilder, M.; Soliman, S.; Meyrignac, O.; Talabard, M.P.; 581
    Lamarque, J.P.; et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID- 582
    19 patients. *Nat. Commun.* **2021**, *12*, 1–11, doi:10.1038/s41467-020-20657-4. 583

49. Saood, A.; Hatem, I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. *BMC Med.* 584
    Imaging **2021**, *21*, 1–10, doi:10.1186/s12880-020-00529-5. 585

50. Nguyen-Phuoc, T.; Li, C.; Theis, L.; Richardt, C.; Yang, Y.L. HoloGAN: Unsupervised learning of 3D representations from 586
    natural images. *Proc. - 2019 Int. Conf. Comput. Vis. Work. ICCVW 2019* **2019**, 2037–2040, doi:10.1109/ICCVW.2019.00255. 587

51. Öngün, C.; Temizel, A. Paired 3D model generation with conditional generative adversarial networks. *Lect. Notes Comput.* 588
    Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) **2019**, *11129 LNCS*, 473–487, doi:10.1007/978-3-030- 589
    11009-3_29. 590

591