The use of deep learning to automate the segmentation of the skeleton from CT volumes of pigs

Johannes Kvam,†,‡ Lars Erik Gangsei,§ Jørgen Kongsro,† and Anne H. Schistad Solberg¶

†Norsvin SA, Hamar, Norway; ‡Department of Circulation and Medical Imaging, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; §Animalia, Oslo, Norway; ¶Norwegian University of Life Sciences (NMBU), Department of Biostatistics, Ås, Norway; and Department of Informatics, University of Oslo, Oslo, Norway

ABSTRACT: Computed tomography (CT) scanning of pigs has been shown to produce detailed phenotypes useful in pig breeding. Due to the large number of individuals scanned and corresponding large data sets, there is a need for automatic tools for analysis of these data sets. In this paper, the feasibility of deep learning for fully automatic segmentation of the skeleton of pigs from CT volumes is explored. To maximize performance, given the training data available, a series of problem simplifications are applied. The deep-learning approach can replace our currently used semiautomatic solution, with increased robustness and little or no need for manual control. Accuracy was highly affected by training data, and expanding the training set can further increase performance making this approach especially promising.

Key words: computed tomography, deep learning, image analysis, pigs

INTRODUCTION

Segmentation of bodies using noninvasive imaging methods such as computed tomography (CT), ultrasound (UL), or magnetic resonance imaging (MRI) is of great importance in medicine, biology, and animal science. These methods have given us the possibility to measure bodies, detect, diagnose, and treat diseases in a noninvasive way. The size and scale of such data sets during the last years have given rise to a demand for more automated interpretation and analysis, compared with interpretation performed by experts such as radiologists, researchers, doctors, or veterinarians. While there is still a gap between advances in medical imaging technologies and computational medical analysis, this gap has recently started narrowing with the help of machine learning techniques (Shen et al., 2017). Machine learning provides an effective way to automate the analysis and diagnosis for medical images (Wang and Summers, 2012).

In pig breeding and genetics, CT has been used since the early eighties (Skjervold et al., 1981). Single slice images from different locations were used in order to predict the in vivo body composition and fatness of the pig. Since then, developments toward whole body helical scanning have improved the accuracy in terms of body composition traits (Gjerlaug-Enger et al., 2012).
Helical scanning makes it possible to generate large amount of information, and as a result, large data sets of the object being scanned. In order to automate and sample more detailed phenotypes used in genetics, an atlas segmentation approach was introduced (Gangsei et al., 2016). In vivo atlas segmentation of CT images, and associated methodology, has proven to be an effective method for registration of different parts of the pig, i.e., like the composition of meat cuts, meat quality, and diagnostics of diseases. Traits like relative size of and leanness of commercial cuts are heritable (Kongsro et al., 2017), thus atlas segmentation results are huge assets in a breeding perspective. Nordbo et al. (2018) developed a number of traits describing the morphology of the shoulder blade. Shoulder blades were segmented (Gangsei and Kongsro, 2016) in silico from CT images of test boars. They found moderate-to-high heritability of the morphological traits which were also genetically correlated to shoulder lesions. The atlas segmentation is based on landmarks on the surface of the pig skeleton, which is segmented by applying a simple threshold at 200 Hounsfield units (HU). Landmarks are set for each of the larger skeletal structures. Until now, these structures have been identified by a version of the method specified in Gangsei and Kongsro (2016). We refer to this method as the reference method throughout the paper. The reference method fails to do a correct segmentation in a substantial proportion of cases and require manual intervention and quality control in most cases. The need for a more automated method for segmenting bones in the pig skeleton is of great importance in order to apply atlas-based segmentation on a larger scale in a commercial breeding program of pigs. Machine learning, and in particular deep learning, has been shown to be a superior method for segmentation and classification of images in medical images (Kumar et al., 2015; Roth et al., 2015; Cheng et al., 2016). Cheng et al. (2016) even argue that deep-learning techniques might potentially change the design paradigm of the computer-aided diagnostic systems.

The aim of this study was to investigate the feasibility of deep learning as a method for segmentation and classification of different parts of the skeleton in CT volumes of pigs.

**MATERIALS AND METHODS**

Deep learning is a branch of machine learning that has been revitalized in recent years due to its performance in image analysis (Krizhevsky et al., 2012). The term deep learning heirs from the fact that deep convolutional neural networks (CNNs) are trained to learn rich hierarchical feature sets (LeCun et al., 2015). These networks are composed of successive convolution operations which are insensitive to the spatial locality of the features meaning that the same features can be identified in multiple locations in images. The depth of the CNN determines the complexity of the features the network is able to recognize given enough training examples are available to tune the network.

To utilize deep learning, a large amount of training examples are needed in order to train all parts of the network. Exactly how much training data are needed is currently an unanswered research question. Elements such as diversity of features, network architecture, and overall problem complexity can vary greatly between both problems and solutions. For supervised learning (LeCun et al., 2015), problems are usually either formulated as classification or regression problems. In classification problems, we categorize the content of images into a discrete set of predefined categories; e.g., determining whether class A or B is in the image. Segmentation is hence typically formulated as a classification problem. For regression problems, the CNN produces a continuous output; e.g., how many instances of class A are present. For 2D images, well-known deep-learning methods are available, but for 3D segmentation, it is still a difficult task for machine-learning or deep CNNs to segment structures from medical images due to several mutually affected challenges. These challenges include complicated anatomical environments in volumetric images, optimization difficulties of 3D networks, and inadequacy of training samples (Dou et al., 2017). In the following sections, we will describe how we perform full-volume segmentation and classification of CT volumes from pigs. By using 2D projections and successively simplifying the images, we are able to utilize proven 2D deep-learning methods to produce 3D segmentations of large full body scans.

The code used in this paper has been inspired by these examples: https://github.com/leriomaggio/deep-learning-keras-tensorflow; https://blog.keras.io/category/tutorials.html.

**3D Through 2D Projections**

In contrast to human medicine, the individuals being scanned are not well behaved. Due to the fact that these are farmed animals to be used for food shortly after CT scanning, the animals are not fixed, and sedation is used instead of anesthesia due
to a shorter withdrawal time. This can cause limbs to be entangled in a variety of ways which affects image quality, making interpretation of the images a challenge. As we want to investigate the feasibility of a fully automated machine learning–driven atlas segmentation, we attempt to reduce the problem complexity by splitting the overall problem into less confounded subtasks. Prior knowledge of pig anatomy and characteristics of in vivo CT scans of pigs is extensively utilized for creating a well adapted and suitable series of subproblems. The overall goal being to maximize the ratio of accuracy to the number of training samples available. As 3D CNNs are still a challenge, particularly in large volumes, we chose to reformulate the problem into a series of 2D problems combined to form a 3D volume. To produce full-volume segmentation, we rely on successive 2D segmentations in the coronal and sagittal images which are then combined to produce a 3D “region of interest” (ROI) (Martin and Aggarwal, 1983). This approach is illustrated in Figure 1. The input coronal and sagittal images are of size 512 × 1250. Their pixel values are non-negative integers reflecting the sum of bone voxels in the associated projection of the binary 3D input mask size 512 × 512 × 1250.

The sagittal and coronal images possess different characteristics. For instance, the forelimbs and hindlimbs are in general recognizable in both sagittal and coronal images, especially if the spine is removed. However, left- and right-side limbs are indistinguishable in the sagittal image, also for the human eye, but distinguishable in the coronal image. As the 2D projections can be cluttered due to overlapping structures, the segmented regions are removed from the CT volume to produce less confounded coronal and sagittal input images for subsequent tasks.

**Task Description**

The final output of the segmentation algorithm is a skeleton which is segmented in 24 structures. Five of these structures, *cranium, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, and sacrum + coccyx*, constitute what we will refer to as the spine, i.e., the major part of the axial skeleton. The remaining part of the axial skeleton contains three major structures: *sternum, left and right costae*. The appendicular skeleton contains 16 major structures, i.e., left and right side of *pelvis, femur, tibia + fibula, tarsal + metatarsal, scapula, humerus, radius + ulna, and carpal + metacarpal*. Furthermore, the number of *thoracic vertebrae, lumbar vertebrae, and costae*, which vary between pigs, is predicted.

The full process is divided into five consecutive and interdependent subproblems as shown in Figure 2. The first step is to segment and remove the spine from the CT volume. The composition of the spine is further classified by a separate CNN, while the remaining part of the skeleton is the input to a third network that identifies the limbs and *sternum*. The segmented limbs are removed from the volume, and the limbs on the left and right side are individually classified using a fourth network. Finally, the remaining skeleton consisting of the left and right side of the ribcage is classified by the fifth CNN. In the following paragraphs, a more detailed

![Figure 1. Combination of sagittal and coronal projections to produce a 3D volume segmentation.](image)
A description of each task is given. For each task, we train a separate CNN, with the same architecture described in section CNN Architecture. An overview of the input, output predictions, and training data for each CNN is given in Table 1.

**Segmentation of the spine.** Based on the sagittal and coronal image of the full skeleton, the CNN was trained to segment one single mask containing the spine. By applying the principles of section 3D Through 2D Projections, the spine was segmented from the remaining part of the skeleton. A binary mask separating the spine from the remaining skeleton constitutes the output from the spine-segmentation task. The centerline of the spine is utilized to partition the left- and right-hand side costae, see section Segmentation of limbs. Spine segmentation is the initial task, thus its stability and precision crucial for the performance of subsequent tasks, which is heavily influenced by errors in the spine segmentation.

**Classification of the spine.** Using the spine mask from section Segmentation of the spine, the coronal and sagittal images of isolated spine are used as input for classification of five well-defined anatomical classes: cranium, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, sacrum + coccyx. In addition, a regression network is added for prediction of the number of thoracic and lumbar vertebrae. These can vary substantially between individuals opposed to the number of cervical vertebrae which is approximately constant ($N = 7$) (King and Roberts, 1960). This CNN structure combining five anatomical classes and regression is simpler than incorporating one mask per vertebra,

![Diagram](image-url)
a structure with ≈33 classes. The information benefit from this simpler structure is close to equal to the more complex structure, due to the extra information from the regression part. All output, i.e., the five anatomical classes and vertebra numbers, are part of the final result.

**Segmentation of limbs.** The animals are not fixated and hence rarely outstretched while scanned, consequently identifying the individual bones in forelimbs and hindlimbs is challenging in the coronal view. Identification of individual bones is easier in the sagittal view. However, determining whether the bone in question belongs to either the left or right limb is not possible relying on the sagittal view alone. This is illustrated in Figure 3, where a t-distributed stochastic neighbor embedding approach is utilized to evaluate class separability in the coronal and sagittal views. In order to simplify the classification, we introduce an intermediate step by training an additional CNN which identifies the left and right limbs. This allows us to split the volume in a left and right side for which individual bones are classified independently by

| CNN                        | Input                      | Problem          | Prediction                  | Training data No. individuals |
|-----------------------------|----------------------------|------------------|-----------------------------|-------------------------------|
| #1 Spine segmentation      | Whole skeleton             | Classification   | Spine                       | 1,735                         |
| #2 Spine classification    | Spine                      | Classification   | Cranium                     | 1,000                         |
|                            |                            |                  | Cervical V.                 |                               |
|                            |                            |                  | Thoracic V.                 |                               |
|                            |                            |                  | Lumbar V.                   |                               |
|                            |                            |                  | Sacrum + Coccyx             |                               |
|                            |                            | Regression       | Vertebras                   |                               |
|                            |                            |                  | Lumbar vertebrae            |                               |
| #3 Limb segmentation       | Skeleton without spine     | Classification   | Left forelimb               | 1,000                         |
|                            |                            |                  | Right forelimb              |                               |
|                            |                            |                  | Left hindlimb               |                               |
|                            |                            |                  | Right hindlimb              |                               |
|                            |                            |                  | Sternum                     |                               |
| #4 Limb classification     | Isolated left and right limbs | Classification | Pelvis                      | 857                           |
|                            |                            |                  | Femur                       |                               |
|                            |                            |                  | Tibia + Fibula              |                               |
|                            |                            |                  | Tarsal + Metatarsal         |                               |
|                            |                            |                  | Scapula                     |                               |
|                            |                            |                  | Humerus                     |                               |
|                            |                            |                  | Radius + Ulna               |                               |
|                            |                            |                  | Carpal + Metacarpal         |                               |
| #5 Rib segmentation        | Isolated left and right side of the ribcage | Classification | Costae                      | 500                           |
|                            |                            | Regression       | Number of costae            |                               |

*Figure 3.* Visualization of class separation using t-distributed stochastic neighbor embedding. The classes shown are for the limb-segmentation task. Left and right limbs are distinguishable in the coronal view but indistinguishable in the sagittal view.
the subsequent operation described in section Classification of limbs.

The input for this task is the skeleton excluding the already segmented spine, see section Segmentation of the spine. The CNN segments the sternum, which is part of the final result, and identifies the four main limbs, i.e., two forelimbs and two hindlimbs with a left and right side, respectively, which are used as input in subsequent tasks. Thus, five classes are implemented in the CNN, reflecting the sternum and the four limbs. Based on the masks from the CNN, a series of operations were used to produce the final output. First, the sternum was segmented by the principles described in section 3D Through 2D Projections. After removal of the sternum, 3D masks containing the left- and right-side limbs were constructed by the same methodology. As expected a priori, the performance of the CNN for distinguishing between the left and right side of the animal in the sagittal view was poor. Thus, the sagittal CNN output masks identifying left and right side were combined. The combined mask in the sagittal view, in addition to the side information from coronal view, was sufficient to segment the limbs and decide which side they belonged to. Finally, after removal of the limbs, the remaining skeleton, where the costae constitutes the bulk, was split in a left and right side by utilizing the centerlines of the sternum and spine. Segmentation of limbs is a key task in the process as it yields both the final segmentation of the sternum and masks for the limbs used in the final two tasks.

Classification of limbs. The isolated left and right limbs, segmented in section Segmentation of limbs, constituted the input for the limb classification task. The classification of limbs yields the final segmentation for eight classes: pelvis, femur, tibia + fibula, tarsal + metatarsal, scapula, humerus, radius + ulna and carpal + metacarpal for the left- and right-hand side of the animal. Thus, making it the CNN with the highest number of output classes. When evaluating the performance of the network emphasis was allocated to the sagittal segmentations, as the eight classes in general have considerably smaller overlap in the sagittal compared with the coronal view.

Segmentation of costae (ribs). Using the negated mask produced in limb segmentation, section Segmentation of limbs, the remaining part of the skeleton consisting mainly of the ribcage is produced. The left and right side are represented as individual images and used as input for the costae segmentation task. The output from the CNN is a binary mask which segment costae from the background and smaller remaining objects with HU intensities in the bone range. Furthermore, as the number of costae vary between pigs, a regression network predicting this number is added. All output, i.e., the anatomical class and number of costae, are part of the final result.

CNN Architecture

As the images all arise from the same volumetric data set, the images and features are similar. The tasks involve mainly semantic segmentation and regression. For segmentation, an U-Net architecture was chosen, which is a CNN designed for semantic segmentation. The U-Net concatenates feature at different scales through downsampling and upsampling which has been shown to be beneficial in medical images (Ronneberger et al., 2015). As our training data are limited, we also implemented dropout at each of the scale representations to increase robustness and avoid overfitting (Srivastava et al., 2014). In two of the tasks, we require in addition a continuous output describing how many instances of a particular class are present in the image. In the task of classifying the spine, we need to know the total number of vertebrae and the number of lumbar vertebrae as described in section Classification of the spine. We also need to count the number of costae as described in section Segmentation of costae (ribs) To accomplish these tasks, we add an additional regression network at the end of the classification network in these two cases. This regression network is a CNN designed by the authors consisting of a series of dilated convolutions (Yu and Koltun, 2015) and max-pooling operations to increase field-of-view rapidly. Finally, the features are flattened, and a fully connected layer outputs the estimated number of class instances. An illustration of the architecture of the classification and regression network is shown in Figure 4. For each of the tasks described in section Task Description, we train a separate instance of this CNN with or without the regression network depending on the task.

Loss functions. Choosing the right loss function determines the efficiency of the training process. For semantic segmentation, categorical cross-entropy loss or dice losses (Sudre et al., 2017) are typical loss formulations. The categorical cross-entropy loss maximizes the chance of a correct classification by minimizing the probability that it is mistaken for any other class. In the case of overlapping
classes, as is present in CT images of entangled limbs, these special cases have to be unique classes to be correctly classified. For instance, if the left and right limb can overlap, there has to be a separate overlap class for categorical cross-entropy to be able to robustly classify it. As we are systematically attempting to reduce the number of classes due to limited training data, a dice loss is chosen as it works well for segmentation tasks and does not need explicit handling of overlap between classes. The dice loss measures ratio between the intersect and union of the predictions and ground truth labels: for multiple labels, the generalized dice loss (GDL) is described as follows:

\[
GDL = \frac{\epsilon + 2 \sum_{i=1}^{N} w_i \sum_{l=1}^{L} y_{l,i} \hat{y}_{l,i}}{\epsilon + \sum_{i=1}^{N} w_i \sum_{l=1}^{L} y_{l,i}^{2} + \sum_{l=1}^{L} \hat{y}_{l,i}^{2}}
\]

(1)

where \(y_{l,i}\) and \(\hat{y}_{l,i}\) are the binary ground truth and predicted label of class \(l\), respectively, and \(\epsilon = 1\) is a smoothing term. To avoid class imbalances, each class is weighted by the class weights \(w_i\). Compared with Sudre et al. (2017), we have removed the bias term, and the peak classification score is found at GDL = -1. The class weights were chosen as \(w_i\), which causes the classes to be weighted equally. For binary classification tasks, e.g., classification of spine and ribs, the term cancels and we get the binary dice loss. An interesting variation of the dice loss called the Wasserstein loss described by Fidon et al. (2017) introduces a label distance matrix that punishes some misclassifications more than others. As there is consistency in the relative positioning of some labels in our data set, the Wasserstein loss would be beneficial to avoid crucial misclassifications such as mistaking left and right or front and back. However, due to the different perspective of the sagittal and coronal views described in section 3D Through 2D Projections, the label distance matrix would have to take on different values in the two views. This would require nondifferentiable operations in choosing the correct distance matrix given the particular view. For classification tasks, the generalized dice loss in equation (1) was chosen as the loss function. For the regression tasks involved in the counting of vertebrae and costae, another loss is introduced to train this part of the network. A typical regression loss is the mean square error (MSE) given as follows:

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (r_i - \hat{r}_i)^2
\]

(2)

where \(r_i\) and \(\hat{r}_i\) are the ground truth and predictions, respectively. This loss was chosen due to its simplicity and smooth derivative.

**EXPERIMENTAL SETUP**

Purebred Duroc and Landrace boars from the boar testing station, Norsvin Delta in Norway, were
CT scanned as part of the Topigs Norsvin commercial genetic program. The pigs were CT scanned using a GE Healthcare VCT 32 scanner at 120 kg BW. The protocol and setting used are described by Kongsro et al. (2017). All animals were cared for according to the laws and regulations for keeping pigs in Norway (Regulation for the keeping of 299 pigs in Norway 2003/02/18/175, 2003; Animal Welfare Act 2009/06/19/97, 300 2009).

Data Annotation

In order to annotate training data, results from the reference method were a major asset. The reference method is, as already mentioned, a heuristic, 3D based, and manually controlled technique. The reference method produces masks for all the 24 major bone structures described in section Task Description. Input images for CNNs were constructed for all five subtasks based on 3D bone structure masks. The bone structures in question were composed of the associated individual bones from the reference method. For a CNN with C classes, the corresponding annotated coronal and sagittal masks had a size of $512 \times 1250 \times C$, i.e., C slices of binary masks, each representing one class. The segmentations from the reference method were used to construct these binary masks, where the pixel values in each of the C mask slices were set to one if the pixel contained the actual projection of class in question. All input images and masks, in both training and test data, were manually corrected.

Training, Validation, and Testing

As the annotation process for the different tasks had a varying degree of complexity, the training data for each task were different. For spine segmentation, a total of 3,470 = 1,735 sagittal + 1,735 coronal images from 1,735 individuals were used for training. For the other tasks, 2,000 images were used for training the networks. For the tasks of spine classification and limb segmentation, the 2,000 images were based on 1,000 individuals. For limb classification and costae segmentation, the number of individuals was approximately 500 as each individual has four associated images as the left and right side are separated. After the first round of training, we reached performances in line with our requirements, except for the limb classification task (section Classification of limbs). Thus, an additional set of training data, 1,428 images from 357 individuals, was added for a second round of training for this task. In addition, to mimic additional unique individuals, we introduced augmentation of the training data. Randomized deformations weights were applied to a uniform B-spline grid (Rueckert et al., 1999) for each input image and its corresponding labels. Out of the available samples used for training, 90% were used for updating the weights of the CNN. The remaining 10% was used as a separate validation set to monitor the training process and prevent overfitting. The final test set was composed of 500 previously unseen CT volumes.

Implementation. The networks were implemented in Python 3.5.2 (Python Software Foundation, https://www.python.org/) using Keras 2.1.1 (Chollet, 2015) with the TensorFlow backend and trained on a NVIDIA GTX 1080 TI GPU using the Adam optimizer. Images were normalized to $\mu = 0$ and $\sigma = 1$. Due to the large size of the images ($512 \times 1250$), the batch size was kept small due to memory restrictions, typically 3–6 images. The small batch size increases training time but with a potential benefit to accuracy due to the stochastic nature of the training process. For tasks involving both classification and regression, the classifier was trained first and kept constant due to the stochastic nature of the training process. For training the regression network. Training was stopped when the validation loss stopped improving. As an example, the training curves for spine classification task are shown in Figure 5.

Performance Evaluation

The different tasks done by the neural networks involved a combination of classification and regression tasks. For classification tasks, the mean dice score was used to evaluate performance. As the dice score was used to evaluate performance. As the dice

$$\overline{DS} = \frac{2 \sum_{i=1}^{N} \sum_{j=1}^{N_{l}} N_{y_{ij}} \sum_{j=1}^{N} y_{ij} \tilde{y}_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N_{l}} N_{y_{ij}} \left( \sum_{j=1}^{N} y_{ij} + \sum_{j=1}^{N} \tilde{y}_{ij} \right)}$$

Figure 5. Training curves for classification and regression for the spine classification task. Whole lines show the training loss, and dashed lines show the validation loss. An epoch is a measure of the number of times all the training examples have been used to update the weights of the network.
scores (DS) are not necessarily symmetrically distributed around their mean, we included accuracy as an additional measurement for the performance of the CNNs, identifying outliers, i.e., failed classifications. Accuracy was calculated by the number of individuals with $\overline{DS} > D_T$ relative to the population. The threshold was chosen through visual inspection, where $D_T = 0.95$ was chosen as it yielded satisfactory results in all segmentation tasks. For regression tasks, the performance was evaluated using the “mean-square-error” described in equation (2). For calculation of accuracy, the decimal output was first rounded to the closest integer. The accuracy was calculated as the proportion of individuals where there were no errors in the rounded prediction relative to the test population.

**RESULTS AND DISCUSSION**

In this section, the performance of each individual task on the $N = 500$ test individuals is presented and discussed. Finally, we discuss some of the implications of the overall results.

**Segmentation of the Spine**

As described in section Segmentation of the spine, the main purpose of the spine segmentation is to simplify the CT volume in the following steps by effectively dividing the volume into a left and right side. Out of all the tasks, this had the most training data and the lowest number of classes, which made us expect good performance. The mean dice score over the $N = 500$ test individuals was $\overline{DS} > 0.99$ both in the coronal and sagittal view. Manual control deemed all segmentations to be satisfactory.

**Classification of the Spine**

In Table 2, the classification and regression results are shown for classification of the spine. With the chosen categorization of vertebrae as described in section Classification of the spine, we achieved a ≥97% accuracy on all labels in the coronal view and a >95% accuracy in the sagittal view. Similarly, for regression tasks, there was a discrepancy in the performance in the coronal and sagittal view. Visually, it is easier to count the total number of vertebrae in the sagittal view as the transition between vertebrae is clearer. However, the thoracic and lumbar vertebrae are more distinguishable in the coronal view due to the more characteristic transverse process. As mentioned in section Classification of the spine, the number of cervical vertebrae is approximately constant, and the number of thoracic vertebrae is given as $N_{t,v} = N_v - N_{l,v} - 7$.

**Segmentation of Limbs**

The performance of the limbs classifier is shown in Table 3. As we can see, there is a drastic difference in the performance in the sagittal and coronal view. As discussed in section Segmentation of limbs, this is expected as there is simply no information in the sagittal 2D projection of the CT volume that allows distinguishing between left and right, this was further illustrated in Figure 3. This shows that the CNN is not able to distinguish what the human eye cannot. However, as also pointed out, producing a 3D segmentation can still be facilitated by combining left and right masks in the sagittal view. Through combination of classes in the sagittal view, the overall accuracy is ~95% which is satisfactory.

**Classification of Limbs**

In the generation of training data as described in section Training, Validation, and Testing, the generated inputs and outputs were manually controlled and cleaned of everything but the structures of interest. However, as this task is dependent on the performance of the spine segmentation and limb

| Class                | Sagittal DS | Sagittal Acc. % | Coronal DS | Coronal Acc. % |
|----------------------|-------------|-----------------|------------|----------------|
| Cranium              | 1.000       | 100.0           | 1.000      | 99.8           |
| Cervical vertebrae    | 0.996       | 99.6            | 1.000      | 99.8           |
| Thoracic vertebrae    | 0.994       | 94.2            | 0.996      | 97.0           |
| Lumbar vertebrae      | 0.997       | 98.2            | 0.996      | 97.4           |
| Sacrum + Coccyx       | 0.997       | 98.2            | 0.998      | 99.8           |
| N                     | MSE         | Acc. %          | MSE        | Acc. %         |
| Vertebras             | 0.045       | 97.4            | 0.052      | 97.0           |
| Lumbar vertebrae      | 0.074       | 93.2            | 0.045      | 95.8           |

| Class                | Sagittal DS | Sagittal Acc. % | Coronal DS | Coronal Acc. % |
|----------------------|-------------|-----------------|------------|----------------|
| Left forelimb         | 0.843       | 4.0             | 0.992      | 97.8           |
| Right forelimb        | 0.850       | 4.0             | 0.993      | 98.2           |
| Left hindlimb         | 0.770       | 2.4             | 0.990      | 96.8           |
| Right hindlimb        | 0.780       | 2.0             | 0.990      | 95.8           |
| Sternum               | 0.967       | 84.9            | 0.983      | 94.6           |
segmentation, CNN-compounded errors were present in the input data set. These errors manifested themselves in two ways. The first source of error was due to misclassifications or partial failures in the segmentation. Due to the high performance of the spine segmentation CNN, these types of errors mainly originated from the CNN performing limb segmentation. These failures limit the maximum achievable dice scores as small parts of the input data can be missing. This indicates that future efforts to improve performance should be focused on improving the limb-segmentation network. The other source of compounded errors, which proved to be the most challenging, was noise-like artefacts in the input image due to losses in the 3D segmentation and corresponding 2D projections. These types of artefacts were not present in the training data, and it became a major source for misclassifications. These artefacts could have been removed using traditional preprocessing image filtering techniques, but to make the CNN robust to these types of errors, we conducted another round of training with a training set that contained the same types of artefacts. The expanded training set increased the average dice coefficient, DS, by between 0.05 and 0.3 which translated into approximately a corresponding increase in accuracy of 34%–95%. The final classification scores are shown in Table 4. The results confirm that identifying the individual bones is more challenging in the coronal view as discussed in section Classification of Limbs. In the sagittal view, the results were satisfactory for most classes; however, some manual correction of the carpal + metacarpal and radius + ulna might be necessary if deviating atlas segmentation results are encountered.

Segmentation of Ribs

Segmentation of the ribs achieved acceptable performance with a DS >0.99 in both the coronal and sagittal views with an accuracy of >96%. Some compound errors as described in section Classification of Limbs were present; however, the main source of error was so-called half-ribs (Fredeen and Newman, 1962), underdeveloped ribs that are barely visible in the CT image. As they are hard to detect, we suspect that they may not have been consistently labeled in the training data; however, in the test set, they were consistently included. As a consequence, the performance of the regression network is perhaps underestimated as it is dependent upon the labels from the classifier which often fails to correctly classify these types of ribs. Consequently, the accuracy for the regression was 78.5% in the coronal and 89.4% in the sagittal view. Due to the classifiers problem with half-ribs, the bone was in most cases partially counted, identified as a fraction in the output. As these fractions are rounded to the closest integer, small variations can have a large impact on the accuracy calculation. The accuracy estimate is hence pessimistic, and the performance is deemed acceptable for the application.

Implications

The amount of CT data available from individual animals described in this paper is unique. The authors had access to several thousand animals (>20,000) CT scanned in the period of 2012–2018 using the same CT-scanning protocol. Development of methods in order to analyze and model these data is crucial not only for the animal science community but possibly also for the scientific community in general. The problem complexity reduction that has been applied has both upsides and downsides. A major upside was the relatively short training time and accuracy of each individual network given the training data available. This allowed us to evaluate parts of the segmentation process independently. Verifying annotations were also simpler as each label set was sparse and easily separable through visual inspection. However, there are two main downsides to this approach. Firstly, the splitting into subproblems causes a compounding of error due to classification errors in the preceding networks. This is especially apparent in the task of limb classification, where an additional training round was needed to make the network robust to these types of consequential errors. Luckily, the cost of training this robustness was minimal. The other downside is the inefficiency compared with a single network. Due to the current processing chain, complete CT volume segmentation is a sequentially

Table 4. Classification results for classification of limbs

| Class             | Sagittal (DS, Acc. %) | Coronal (DS, Acc. %) |
|-------------------|-----------------------|----------------------|
| Pelvis            | 0.970 (94.4, 0.941)   | 0.966 (92.6, 0.976)  |
| Femur             | 0.981 (97.0, 0.950)   | 0.998 (99.0, 0.975)  |
| Tibia + Fibula    | 0.966 (90.4, 0.950)   | 0.964 (91.8, 0.975)  |
| Tarsal + Metatarsal| 0.972 (90.4, 0.950)   | 0.988 (91.8, 0.975)  |
| Scapula           | 0.957 (83.3, 0.936)   | 0.948 (77.8, 0.950)  |
| Carpal + Metacarpal| 0.948 (77.8, 0.950)  | 0.957 (83.3, 0.936)  |

Translate basic science to industry innovation
dependent computation. This is not a critical element but more a question of elegance.

The computational bottleneck in our currently used semiautomatic solution is the need for manual input from an operator. This limits the throughput of the algorithm. The solution proposed here can, in its current state, be deployed for automatic full-volume landmark detection in atlas segmentation. Manual control for a small subset of individuals with deviating atlas segmentation results might still be needed. In addition, the controlled cases can be reintroduced in training which we saw had a significant impact on the performance on the limb classification task described in section Classification of Limbs. This process allows us to efficiently build up a large database of annotated volumes. Once a large collection of annotated data sets has been built up moving toward a single full-volume segmentation network is a natural next step. Especially, interesting is the combination of a convolutional and recurrent neural network (CNN + RNN) as described in Donahue et al. (2017); Pinheiro and Collobert (2014) for a slice-by-slice full-volume classification.

As CT is known to be robust for segmenting bone from soft tissue and the CNN input data are normalized, we believe that changes in the CT protocol will not have a significant effect on the results in this paper.

CONCLUSION

In this paper, the feasibility of fully automatic deep learning–driven segmentation of different parts of the pig skeleton from volumetric CT data was investigated. To maximize performance, given the training samples available, a series of steps were taken to simplify the problem. The final 2D-based solution can replace our currently utilized 3D-based method but more robust and with little or no need for manual intervention. In addition, accuracy was improved by introducing more training data, confirming the feasibility of the approach.

REFERENCES

Cheng, J. Z., D. Ni, Y. H. Chou, J. Qin, C. M. Tiu, Y. C. Chang, C. S. Huang, D. Shen, and C. M. Chen. 2016. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci. Rep. 6:24454. doi:10.1038/srep24454.

Chollet, F. 2015. Keras. [accessed February 14, 2018] https://github.com/fchollet/keras.

Donahue, J., L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell. 2017. Long-term recurrent convolutional networks for visual recognition and description. IEEE Trans. Pattern Anal. Mach. Intell. 39:677–691. doi:10.1109/TPAMI.2016.2599174

Dou, Q., L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P. A. Heng. 2017. 3D deeply supervised network for automated segmentation of volumetric medical images. Med. Image. Anal. 41:40–54. doi:10.1016/j.media.2017.05.001.

Fidon, L., W. Li, L. C. García-Peraza-Herrera, J. Ekanayake, N. Kitchen, S. Ourselin, and T. Vercauteren. 2018. Generalised Wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks. In: Crimi A., Bakas S., Kuijf H., Menze B., Reyes M., editors. Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries. BrainLes 2017. Springer, Cham: Lecture Notes in Computer Science, vol 10670. doi:10.1007/978-3-319-75238-9_6

Fredeen, H. T., and J. A. Newman. 1962. Rib and vertebral numbers in swine: I. Observation varied in a large population. Can. J. Anim. Sci. 42:232–239. doi:10.4141/cjas62-036

Gangsei, L. E., and J. Kongro. 2016. Automatic segmentation of computed tomography (CT) images of domestic pig skeleton using a 3D expansion of Dijkstra’s algorithm. Comput. Electron. Agric. 121:191–194. doi:10.1016/j.compag.2015.12.002.

Gangsei, L. E., J. Kongro, K. Olsstad, E. Grindflek, and S. Sebo. 2016. Building an in vivo anatomical atlas to close the phenomic gap in animal breeding. Comput. Electron. Agric. 127:739–743. doi:10.1016/j.compag.2016.08.003.

Gjerlaug-Enger, E., J. Kongro, J. Ødegård, L. Aass, and O. Vangen. 2012. Genetic parameters between slaughter pig efficiency and growth rate of different body tissues estimated by computed tomography in live boars of Landrace and Duroc. Animal 6:9–18. doi:10.1017/S1751731111001455.

King, J. W. B., and R. C. Roberts. 1960. Carcass length in the bacon pig: its association with vertebrae numbers and prediction from radiographs of the young pig. Anim. Sci. 2:5965. doi:10.1017/S0003356100033493

Kongsro, J., L. E. Gangsei, T. M. Karlsson-Drangsholt, and E. Grindflek. 2017. Genetic parameters of in vivo primal cuts and body composition (pigatlas) in pigs measured by computed tomography (CT). Trans. Anim. Sci. 1:599–606. doi:10.2527/tas2017.0072

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., C. J. C. Burges, L. Bottou, K. Q. Weinberger, editors. Advances in Neural Information Processing Systems 25. New York: Curran Associates, Inc., p. 1097–1105. doi:10.1017/S0003356100033493

Kumar, D., A. Wong and D. A. Clausi. 2015. Lung nodule classification using deep features in CT images. In: 12th Conference on Computer and Robot Vision. Halifax, NS, Canada; IEEE, p. 133–138. doi:10.1109/CVRV.2015.25

LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. Nature 521:436–440. doi:10.1038/nature14539

Martin, W. N., and J. K. Aggarwal. 1983. Volumetric description of objects from multiple views. IEEE Trans. Pattern Anal. Mach. Intell. 5:150–158. doi:10.1109/TPAMI.1983.476367

Nordbo, Ø., L. E. Gangsei, T. Aasmundstad, E. Grindflek, and J. Kongro. 2018. The genetic correlation between scapula shape and shoulder lesions in sows. J. Anim. Sci. 96:1237–1245. doi:10.1093/jas/sky051
Deep-learning CT segmentation of pig skeleton

Pinheiro, P., R. Collobert. 2014. Recurrent convolutional neural networks for scene labeling. In: Xing, E. P. and T. Jebara, editors. Proceedings of the 31st International Conference on Machine Learning, vol 32, p. 82–90, Beijing, China.

Ronneberger O., P. Fischer, T. Brox. 2015. U-net: convolutional networks for biomedical image segmentation. In: Navab N., Hornegger J., Wells W., Frangi A., editors. Medical image computing and computer-assisted intervention – MICCAI 2015. Lecture Notes in Computer Science, vol. 9351, Cham: Springer; p. 234–241. doi:10.1007/978-3-319-24574-4_28

Roth, H. R., C. T. Lee, H. C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, R. M. Summers. 2015. Anatomy-specific classification of medical images using deep convolutional nets. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI); New York, NY, 2015, p. 101–104. doi:10.1109/ISBI.2015.7163826

Rueckert, D., L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. 1999. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging 18:712–721. doi:10.1109/42.796284

Shen, D., G. Wu, and H. I. Suk. 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19:221–248. doi:10.1146/annurev-bioeng-071516-044442

Skjervold, H., K. Grønseth, O. Vangen, and A. Evensen. 1981. In vivo estimation of body composition by computerized tomography. Z. Tierzücht. 98:77–79. doi:10.1111/j.1439-0388.1981.tb00330.x

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15:1929–1958.

Sudre, C. H., W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso. 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, p. 240–248. Cham: Springer. doi:10.1007/978-3-319-67558-9

Wang, S., and R. M. Summers. 2012. Machine learning and radiology. Med. Image. Anal. 16:933–951. doi:10.1016/j.media.2012.02.005, arXiv:NIHMS150003

Yu, F., and V. Koltun. 2015. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

Translate basic science to industry innovation