Context-Based Segmentation of the Longissimus Muscle in Beef with a Deep Neural Network

Karol Talacha[1], Bartosz Świderski[2], Jarosław Kurek[2], Michał Kruk[2], Andrzej Półtorak[1],
Leszek J. Chmielewski[2], Grzegorz Wieczorek[2], Izabella Antoniuk[2],
Jakub Pach[2], Arkadiusz Orłowski[2]
[1]*Institute of Human Nutrition Sciences*
[2]*Institute of Information Technology*
Warsaw University of Life Sciences – SGGW, Poland

Abstract. The problem of segmenting the cross-section through the *longissimus* muscle in beef carcasses with computer vision methods was investigated. The available data were 111 images of cross-sections coming from 28 cows (typically four images per cow). Training data were the pixels of the muscles, marked manually. The AlexNet deep convolutional neural network was used as the classifier, and single pixels were the classified objects. Each pixel was presented to the network together with its small circular neighbourhood, and with its context represented by the further neighbourhood, darkened by halving the image intensity. The average classification accuracy was 96%. The accuracy without darkening the context was found to be smaller, with a small but statistically significant difference. The segmentation of the *longissimus* muscle is the introductory stage for the next steps of assessing the quality of beef for the alimentary purposes.

Key words: beef carcasses, context-based, segmentation, longissimus muscle, classification, deep convolutional network, beef quality.

1. Introduction

Quality of food is one of the important concerns in contemporary living. Quantities in which meat is produced makes the ocular assessment of various features of meat too time-consuming and prone to subjectivity and errors resulting from fatigue. Computer vision methods are introduced to the meat production industry and research on accurate and effective methods is necessary.

In this paper we shall study the case of segmenting the cross-section of the *longissimus* muscle (called also *longissimus dorsi*). We will consider only the initially prepared transverse sections received during the partition of beef carcasses in which this muscle is the largest connected region, so we shall not go into the anatomic considerations. The knowledge on the location of the *longissimus* will be represented by the images annotated by a human expert.

The segmentation of the images of cross-sections through beef into the region belonging to the muscle of our interest and the other regions is the introductory stage of the analysis aimed at estimating the marbling of this muscle, which is an important factor influencing the quality of meet as an alimentary product. The marbling itself is out of scope of the present paper.

As the classifier we shall use the convolutional neural network (CNN) AlexNet proposed by Alex Krizhevsky et al. in 2017 [9]. Since then this network proved to be successful in many applications.

Survey on the use of computer vision in the meat quality evaluation can be found in [13] from 2004 and in a more recent paper [12] from 2019.

The following papers treat the question of analysis of the images of beef. Jackman, Sun and Allen [7] investigated the *longissimus dorsi* muscle in beef and its marbling and presented a customary system based on clustering and thresholding with contrast enhancement. They have presented very good results of the segmentation in comparison to a method requiring manual intervention of an expert. Agustin and Dijaya [1] recognized the quality and freshness of beef to enable a fast decision to be made at the butcher's. They have shown that a k-nearest neighbour classifier with the colour and textural features is able to recognize the type of meat. Andaya et al. [3] went beyond the muscle segmentation and assessed the meat marbling with a fuzzy logic-based classifier. They have investigated beef as one of the meats.

The research on other kinds of meet, fresh as well as cooked, was the subject of a number of papers. De Guzman et al. [6] assessed pork quality with the support vector machine. Quality of chicken meat was tested with color camera by Barbin et al [4] and it was found theat the results similar to those found with a colorimeter can be found. Color is a very important source of information in meat. Al-Sarayreh, Klette et al. [2] presented interesting results in testing red meat adulteration with hyperspectral imaging. Such imaging was also used by Kamruzzaman et al. [8] to monitor the colour of red meet online. Sun, Young, Liu et al. [11] also studied the relation between the colorimeter results and the colour computer vision results for pork meat.

There are also numerous papers on the recognition of other kinds of food with the image processing methods. For example, Du, Iqbal and Sun [5] analysed the quality of various cooked food.

Finally, convolutional neural networks are used in the domain of meat image analysis. Muñoz, Gou and Fulladosa [10] investigated the contents of intramuscular fat in dry ham slices. Among other things, they pointed to the question of manual labelling the training data which can be problematic.

To our best knowledge, until now AlexNet has not been used to the segmentation of beef, including the *longissimus dorsi*, although many other methods were used in the application to processing the images of meet in general, and beef in particular.

The remaining part of this paper will be organized as follows. In the next Section we shall present the images we have used for this research. In Section 3 the training and classification will be discussed together. Configuration of the network will be shown in

Subsection 3.1. More attention will be paid in Subsection 3.2 to our simple but effective proposition of using the context of a pixel. The accuracy of the results found with the context-based method will be reported in Subsection 3.3, and that received with the typical neighbourhood-based method – in Subsection 3.4. The results will be briefly discussed in Section 4 and the paper will be concluded in Section 5.

2. Images

In the study the images of sections through the body of the cow, with the cross-section through the *musculus longissimus* were analysed. From the images of cross-sections made under various ribs, the subject of our interest were those under ribs 5 and 6, from both sides, which makes four images per cow. There were 28 cows giving 112 images, one image was missing for one cow, so there were 111 images in total for the analysis. Example images from the cow no. 3 are shown in Figs. 1a, b, e and f.

In these images, an expert marked manually the regions belonging to the cross-section through the em musculus longissimus, shown in Figs. 1c, d, g and h. Such marked images, together with their originals, served as data for the training, classification and testing aimed at the segmentation of the images into the two classes.

3. Training and classification

3.1. Configuration of the network

To show the configuration of AlexNet in a detailed and univocal way, the contents of the configuration file is directly presented in Listing 1. The line numbers are shown explicitly. For presentation only, the long lines of text are shown broken, with the continuations deeply indented; in the original file they were long, unbroken lines.

3.2. Presentation of the images and the context

In AlexNet, as the objects for training as well as for classification, the image fragments are shown. At the training stage each fragment is labelled with the class index (muscle: 1, other region: 0). In the classification phase, the image fragment is given at the input, and the index of the class is returned by the network at the output.

For the presentation of the images to the network the concept of context-based recognition was used. We thought of applying something meaningful but simple, going beyond the presentation of a pixel together with its small neighbourhood. As a solution, it was proposed to present the pixel together with the contents of the whole image surrounding it, in which the immediate circular neighbourhood was presented directly, and the remaining part of the image was darkened. In this way, the pixel was shown together with

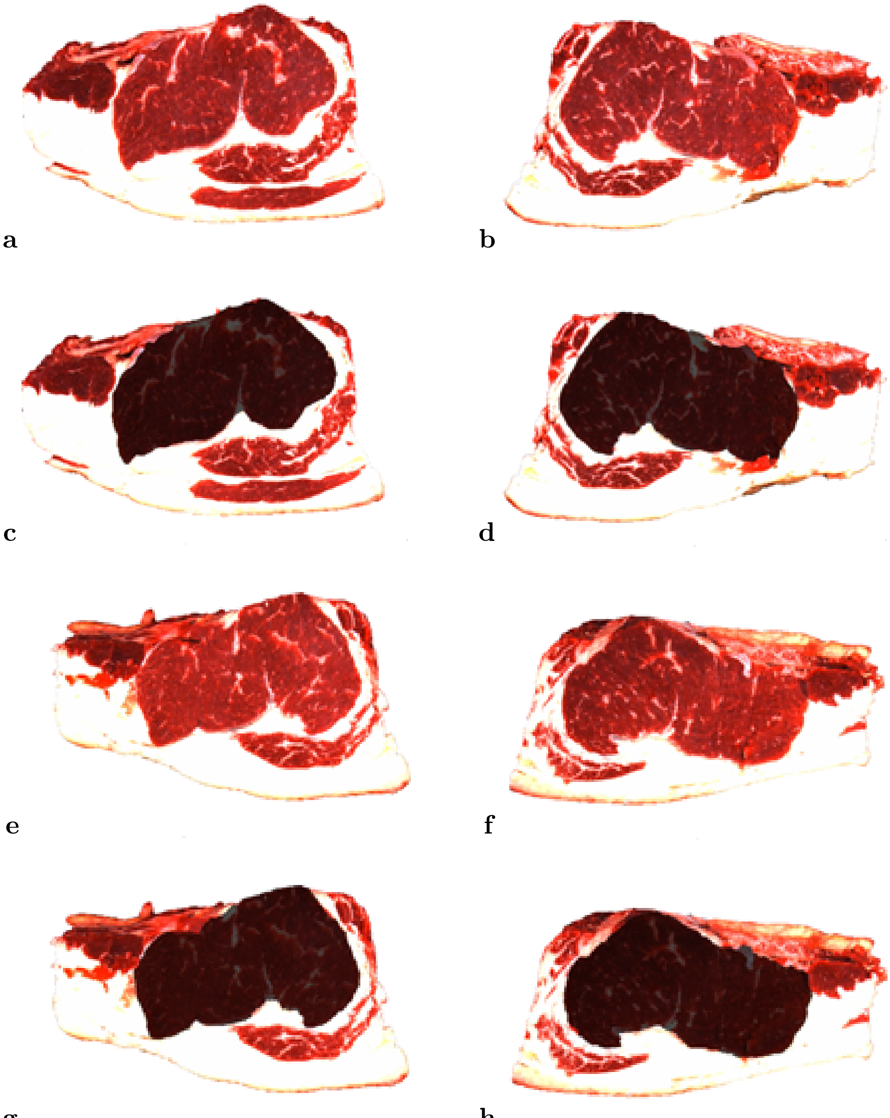

Fig. 1. Images with cross-sections through the *musculus longissimus* and surrounding tissues. **a**, **b**, **e**, **f**: original images; **c**, **d**, **g**, **h**: images as above, with the region of *musculus longissimus* marked in the way of darkening the respective pixels.

```
 1   'data'     Image Input                   227x227x3 images with
                                                 'zerocenter' normalization
 2   'conv1'    Convolution                   96 11x11x3 convolutions with
                                                 stride [4  4] and padding [0
                                                 0]
 3   'relu1'    ReLU                          ReLU
 4   'norm1'    Cross Channel Normalization   cross channel normalization with
                                                 5 channels per element
 5   'pool1'    Max Pooling                   3x3 max pooling with stride [2
                                                 2] and padding [0  0]
 6   'conv2'    Convolution                   256 5x5x48 convolutions with
                                                 stride [1  1] and padding [2
                                                 2]
 7   'relu2'    ReLU                          ReLU
 8   'norm2'    Cross Channel Normalization   cross channel normalization with
                                                 5 channels per element
 9   'pool2'    Max Pooling                   3x3 max pooling with stride [2
                                                 2] and padding [0  0]
10   'conv3'    Convolution                   384 3x3x256 convolutions with
                                                stride [1  1] and padding [1  1]
11   'relu3'    ReLU                          ReLU
12   'conv4'    Convolution                   384 3x3x192 convolutions with
                                                stride [1  1] and padding [1  1]
13   'relu4'    ReLU                          ReLU
14   'conv5'    Convolution                   256 3x3x192 convolutions with
                                                stride [1  1] and padding [1  1]
15   'relu5'    ReLU                          ReLU
16   'pool5'    Max Pooling                   3x3 max pooling with stride [2
                                                2] and padding [0  0]
17   'fc6'      Fully Connected               4096 fully connected layer
18   'relu6'    ReLU                          ReLU
19   'drop6'    Dropout                       50% dropout
20   'fc7'      Fully Connected               4096 fully connected layer
21   'relu7'    ReLU                          ReLU
22   'drop7'    Dropout                       50% dropout
23   'fc8'      Fully Connected               1000 fully connected layer
24   'prob'     Softmax                       softmax
25   'output'   Classification Output         crossentropyex with 'tench',
                                                'goldfish', and 998 other
                                                classes - two classes (pixel
                                                belonging / not belonging to
                                                musculus longissimus)
```

Listing 1. Parameter file for AlexNet. Broken lines shown here as deeply indented.

its immediate neighbours, and the broader context. In detail, the following operations were done for each trained as well as classified pixel.

1. Image was centred at the given pixel.
2. Image around the pixel was trimmed to 227×227 pixels. If the pixel was too close to the image boundary to do this, the missing part was padded with white pixels.
3. Pixels outside the small circular region centred at the pixel, with the area of approx. 1% of the trimmed image area (radius ≈13 pixels), were darkened by reducing the image intensity by a half.

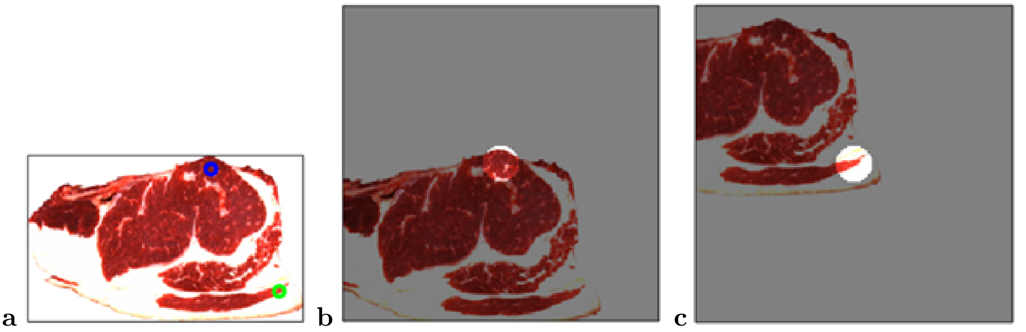

Fig. 2. Image (of Fig. 1a) with two pixels prepared for training or classification. **a**: image with two interesting pixels marked with colours; **b**, **c**: images presented to the network for pixel marked with blue and green, respectively. Grey frames indicate the borders of the pixels.

The examples of these operations are shown in Fig. 2. For the training, the image of Fig. 1b was presented to the network together with the label 1, and the one of Fig. 1c – with the label 0, as an example.

For comparison, the same was done without darkening the context of the pixels. This will be referred to further.

3.3. Classification accuracy

The assessment of the classification accuracy was performed with the cross-validation method. Images from all the cows, except one cow left out for testing, were used to train the classifier, and the results were checked on the images belonging to that testing cow. Such a procedure was repeated with until each cow was the testing one.

The training was a time-consuming task, so for the training phase only a part of pixels of the training images was drawn at random. There were two versions of training: with drawing 3000 pixels for training in each training image, and with drawing 100 pixels for training in each image. In both cases, the testing was performed on all pixels of all images of the testing cow.

The classification quality was measured with the measure of accuracy:

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}\,, \tag{1}$$

where TP, FP, TN and FN are the numbers of true positive, false positive, true negative and false negative classifications of the pixels.

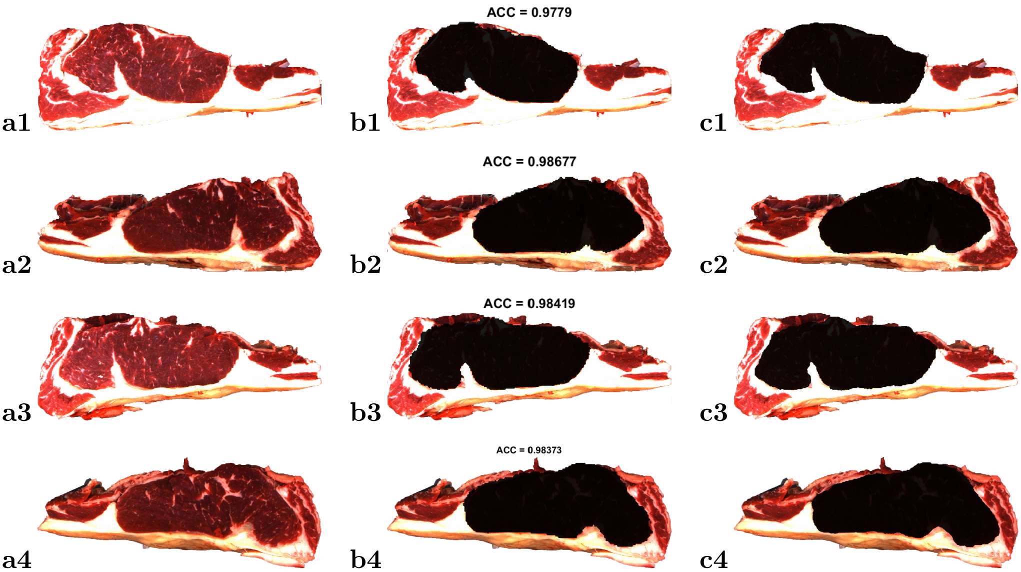

Fig. 3. Examples of results and accuracies found with 3000 training pixels drawn from each image, in a single validation round with just one testing cow (cow no. 1). (**a**) raw image; (**b**) classified image; (**c**) ground truth – training image; (**x1-x4**) subsequent images for this cow.

The example results and accuracies found with 3000 pixels drawn for training from each cow, in a single validation round with just one testing cow, are shown in Fig. 3.

The experiment with all the cows and with drawing 100 pixels from each image for training was performed. To find the overall accuracy, the accuracies were found for each image for each testing cow separately, and averaged for the images. This overall accuracy was found to be 96.27% for 111 images (minimum 93.04%, maximum 98.37%). Selected results are shown in Figs. 4 and 5.

3.4. Accuracy without darkening the context

The same experiment was repeated without darkening the context of the pixels presented to the network. In this case, 100 pixels were drawn from each image for training, as above. The overall accuracy was found to be 95.26% (minimum 91.00%, maximum 97.43%). In 88/111 cases, that is 79%, the results with the darkened context were better than those without darkening. The difference was found to be statistically significant at the p-value much lower than 0.01, which was tested with the *sign test* and *signrank* tests.

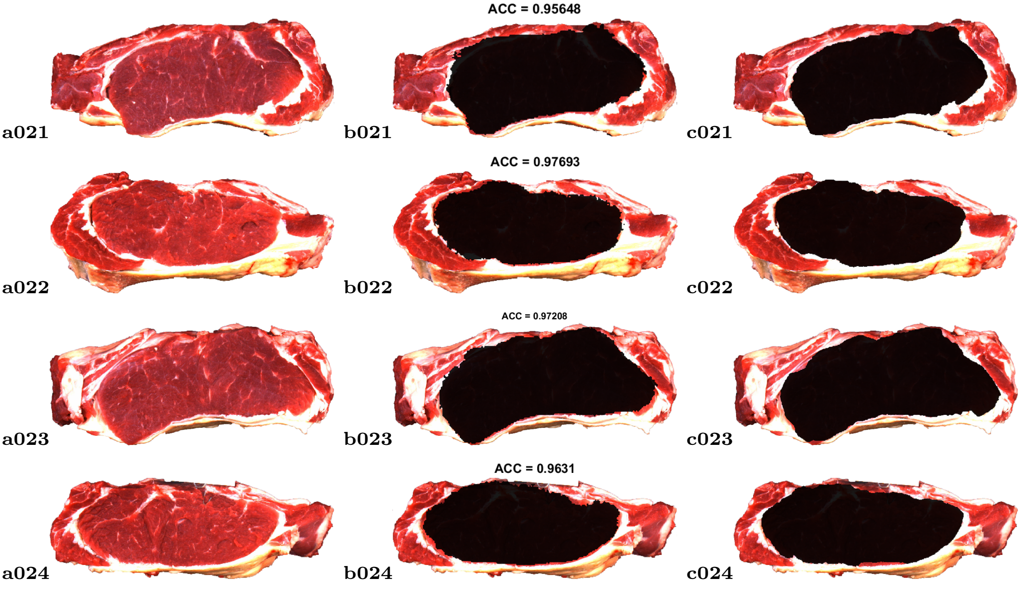

Fig. 4. Examples of results and accuracies found with cross-validation, with 100 training pixels drawn per image, for cow no. 2. (**a**) raw image; (**b**) classified image; (**c**) ground truth – training image; (**xiij**) image for cow *ii*, image *j*.

4. Discussion

The classification method proposed appeared to perform with the average accuracy 96% which can be considered as a promising result. This result was achieved for 100 training pixels per image. Such a small number was a compromise between accuracy and training speed. The accuracy with more training pixels could be larger, which has been confirmed with the experiment with 3000 pixels per image and with a single cross-validation round (one testing cow and 27 training ones), in which the average accuracy was 98.4%.

The speed of classification was limited with the necessity of preparing the neighbourhoods and context for each classified pixel. These operations are simple image shifting and convolutions, and could be performed much more efficiently if dedicated software, or better, dedicated hardware were used.

For the difference in the classification accuracies between the versions of the presentation of pixels to the classifier, with darkened context and with unchanged neighbourhood, it has been demonstrated that the version with darkened context as descried in Section 3.2 was significantly better than that with showing the pixel simply together with its large neighbourhood.

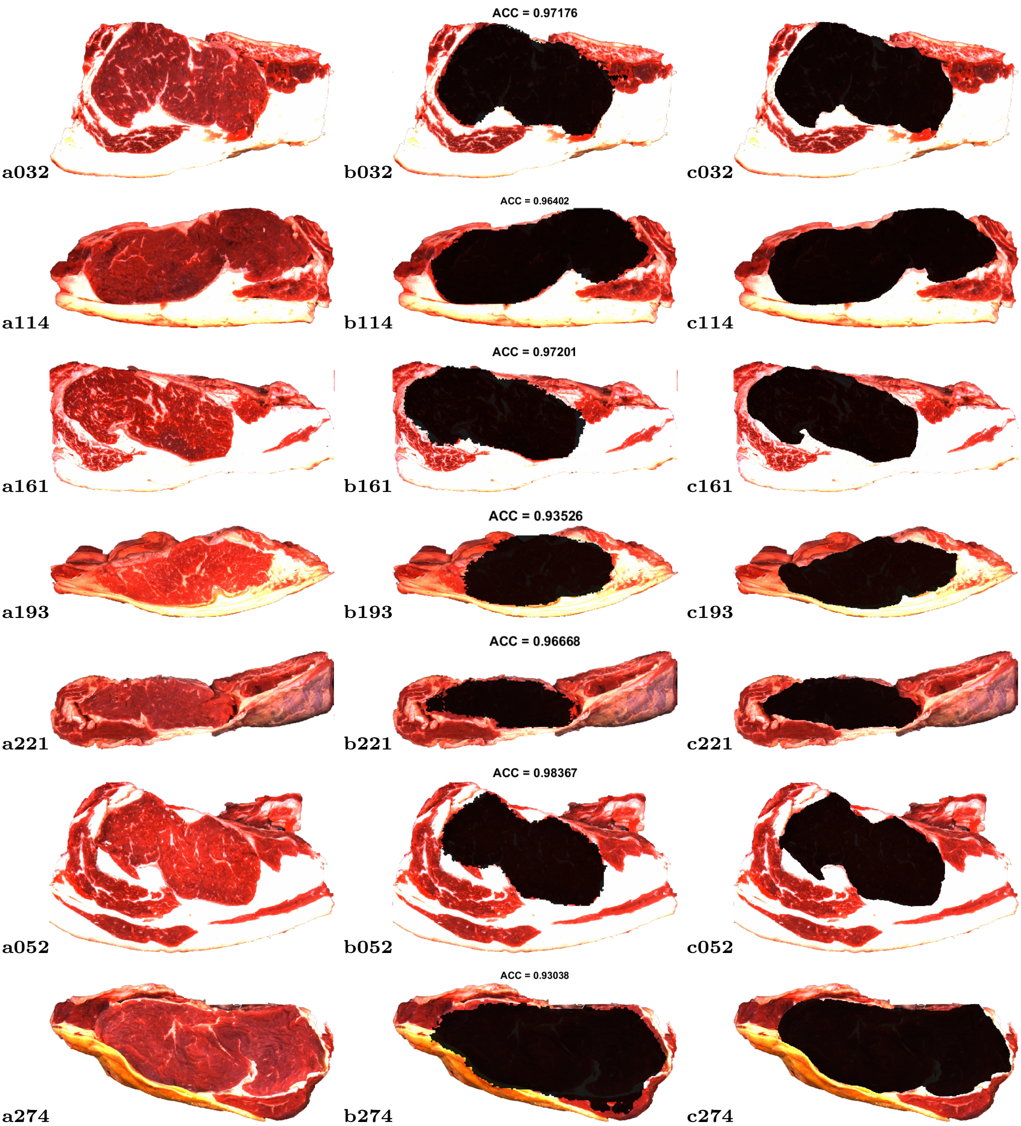

Fig. 5. Examples of results and accuracies found with cross-validation, with 100 training pixels per image, for selected cows and their images. (**a**) raw image; (**b**) classified image; (**c**) ground truth – training image; (**xiij**) image for cow *ii*, image *j*. The last two examples are the extremal cases of accuracy: maximum 0.98367 for cow 5 and minimum 0.93038 for cow 27.

5. Conclusion

To assess the quality of beef carcasses with such measures like, for example, marbling, it is first necessary to find the region occupied by the *longissimus* muscle, which is important from the point of view of cookery. The image segmentation problem stated in this way was the problem of our interest.

As the classifier, the AlexNet deep convolutional neural network was used. As the classified objects, pixels of the images of the cross-sections were considered. The data analysed consisted of 111 images of cross-sections, taken from 28 cows (four images per cow, one image missing). Training data were available for these images in the form of images with pixels belonging to the muscle sought marked manually by human experts. It was proposed to present to the network each pixel together with its small circular neighbourhood, and also together with its context represented by the broader neighbourhood, that is, the further neighbourhood, darkened by reducing the image intensity by a half. The average classification accuracy received with the cross-validation method was at the level of 96% and could be improved to around 98% by performing the training with more data. The accuracy without darkening the context was found to be smaller (95%). The difference was small but statistically significant (p-value < 0.01, for *sign test* and *sigrank* tests).

The proposed methodology, including the context-based classification and application of the deep convolutional network, seems to be a promising possibility in the domain of segmentation of meet images.

References

[1] S. Agustin and R. Dijaya. Beef image classification using k-nearest neighbor algorithm for identification quality and freshness. *Journal of Physics: Conference Series*, 1179:012184, July 2019. doi:10.1088/1742-6596/1179/1/012184.

[2] M. Al-Sarayreh, M. M. Reis, W. Q. Yan, and R. Klette. Detection of red-meat adulteration by deep spectral-spatial features in hyperspectral images. *Journal of Imaging*, 4(5), 2018. doi:10.3390/jimaging4050063.

[3] A. E. Andaya, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa. Meat marbling scoring using image processing with fuzzy logic based classifier. *Int. J. of Scientific & Technology Research*, 8(8):1442–1445, Feb. 2018. `http://www.ijstr.org/research-paper-publishing.php?month=aug2019`.

[4] D. F. Barbin, S. M. Mastelini, S. Barbon, G. F. C. Campos, A. P. A. C. Barbon, and M. Shimokomaki. Digital image analyses as an alternative tool for chicken quality assessment. *Biosystems Engineering*, 144:85–93, 2016. doi:10.1016/j.biosystemseng.2016.01.015.

[5] C.-J. Du, A. Iqbal, and D.-W. Sun. Quality measurement of cooked meats. In D.-W. Sun, editor, *Computer Vision Technology for Food Quality Evaluation (Second Edition)*, chapter 8, pages 195–212. Academic Press, San Diego, second edition, 2016. doi:10.1016/B978-0-12-802232-0.00008-6.

[6] R. J. S. De Guzman, D. N. N. Niro, and A. C. F. Bueno. Pork quality assessment through image segmentation and support vector machine implementation. *Journal of Technology Management and*

Business, 5(2), Feb. 2018. `https://publisher.uthm.edu.my/ojs/index.php/jtmb/article/view/2261`.

[7] P. Jackman, D.-W. Sun, and P. Allen. Automatic segmentation of beef longissimus dorsi muscle and marbling by an adaptable algorithm. *Meat Science*, 83(2):187–194, 2009. doi:10.1016/j.meatsci.2009.03.010.

[8] M. Kamruzzaman, Y. Makino, and S. Oshita. Online monitoring of red meat color using hyperspectral imaging. *Meat Science*, 116:110–117, 2016. doi:10.1016/j.meatsci.2016.02.004.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, May 2017. doi:10.1145/3065386.

[10] I. Muñoz, P. Gou, and E. Fulladosa. Computer image analysis for intramuscular fat segmentation in dry-cured ham slices using convolutional neural networks. *Food Control*, 106:106693, 2019. doi:doi.org/10.1016/j.foodcont.2019.06.019.

[11] X. Sun, J. Young, J. H. Liu, L Bachmeier, R. M. Somers, K. J. Chen, and D. Newman. Prediction of pork color attributes using computer vision system. *Meat Science*, 113:62–64, 2016. doi:10.1016/j.meatsci.2015.11.009.

[12] A. Taheri-Garavand, S. Fatahi, M. Omid, and Y. Makino. Meat quality evaluation based on computer vision technique: A review. *Meat Science*, 156:183–195, 2019. doi:10.1016/j.meatsci.2019.06.002.

[13] J. Tan. Meat quality evaluation by computer vision. *Journal of Food Engineering*, 61(1):27–35, 2004. doi:10.1016/S0260-8774(03)00185-7.