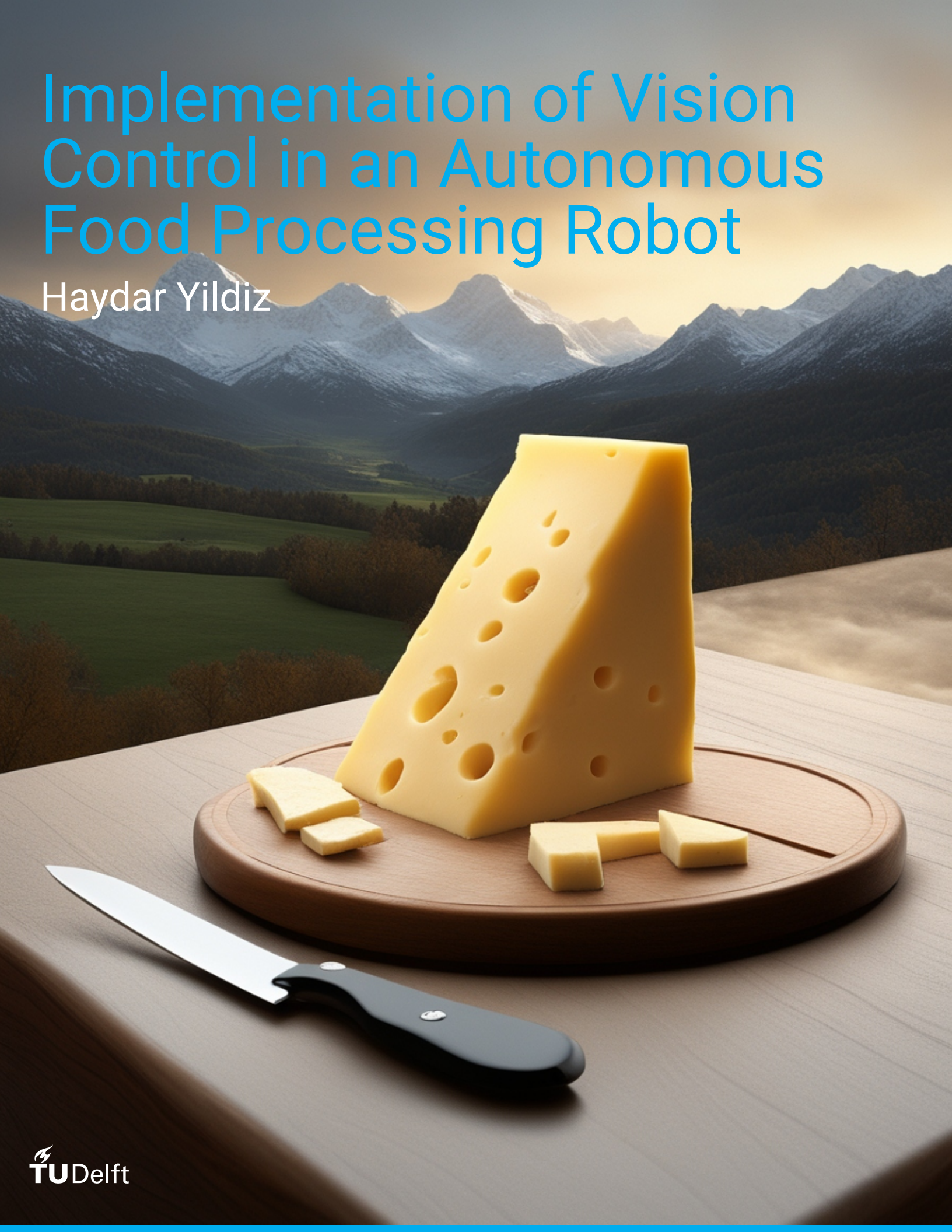
Implementation of Vision Control in an Autonomous Food Processing Robot

Haydar Yildiz



TU Delft

Implementation of Vision Control in an Autonomous Food Processing Robot

by

Haydar Yildiz

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday January 29, 2024 at 10:30 AM.

Student number:	4394666	
Project duration:	February 6, 2023 – January 29, 2024	
Thesis committee:	Dr. ir. Yke Bauke Eisma,	TU Delft, supervisor
	Prof. dr. ir. Martijn Wisse,	TU Delft
	Ir. Stijn Bosma,	Lely, supervisor

TU Delft

Implementation of Vision Control for Curds in Autonomous Cheese Production

***Abstract*— Cheese making is thousands of years old. An important part of the cheese production is the curd size, as this variable has large implications on, among other things, the pH and moisture content of the cheese and the the yield of a cheese in weight. While in traditional cheese making a human expert judges curd size during its cutting process, this is not possible in an autonomous cheese factory and therefore an autonomous method for curd control is desired. As no previous research is done in this area, this study aims to set the ground work for an autonomous method to control curd sizes at the Lely cheese factory. An algorithm is presented to gain insight into the cutting process and to control the curd size during cutting. To assess its practical use at the Lely cheese factory, multiple cheeses are made with varying curd sizes and their moisture content, pH value and cheese yield (in the form of curd fines) are investigated and compared. It was found that adequate control of curd size was possible, with the algorithm performing sufficiently close to the ground truth. The created cheeses at the Lely cheese factory, however, did not show the expected trends at various curd sizes. Recommendations are made to improve the consistency at the Lely cheese factory.**

I. INTRODUCTION

Since 2017 The R&D department of Lely has been working on their next big robot to help the farmer; one that can autonomously make cheese, locally at a farm. This robot closely cooperates with the Lely Astronaut, using its fresh milk to directly create cheese at the farm. The cheese project has, to this date, already produced more than 150 cheeses during testing and is continuously improved until it is ready to present publicly. This research is aimed to make an improvement on the curd control during the cutting process of the cheese robot, to achieve its final goal of autonomously and continuously making cheese.

Cheese General

Cheese is consumed as a complementary part of a main dish, or even enjoyed on its own (Farkye, 2004). It is then not surprising that world production of cheese is more than 18 million tonnes per year and steadily increasing along with the cheese consumption (P. Fox et al., 2017, p. 8). Even more telling about its importance is that around 35% of all milk is turned into cheese (P. Fox et al., 2017, p. 7).

Cheese History

The discovery of cheese started a long time ago, namely more than 7000 years ago (McClure et al., 2018). At that time it was customary to use animal stomachs to transport liquids like milk. When this happens, the (remainders of) rennet in the stomach lead to the coagulation of the milk, which made it possible to extract cheese curds and whey (P. Fox et al., 2017, p. 3), which is believed to have lead to the accidental discovery of cheese (Silanikove et al., 2015). Nowadays there are more than hundreds of cheese sorts being made (P. Fox et al., 2017, p. 2), based on this first discovery and the production processes have evolved during millennia.

Cheese Production

While all those types of cheese are produced slightly differently, the main procedure is the same. In general, the first step it to pasteurize the milk to deactivate pathogens. This can be done by heating the milk for a short time at $\sim 73°C$ or for a longer period at $\sim 64°C$. (McSweeney et al., 2017, p. 42). The next step is the acidification by adding starter culture bacteria to the milk, which produces lactic acid from the present lactose to create the distinctive cheese flavour (McSweeney et al., 2017, p. 201). Next the rennet is added to the milk to coagulate it, forming it into a gel (McSweeney et al., 2017, p. 53). This gel is then cut in smaller pieces named curds. In doing so, the mixture now consist of pieces of curd (solids) and whey (liquid). When the curds are at the desired size, as subjectively determined by the cheesemaker, the mixture is cooked by heating (varying between $\sim 40°C$ and $\sim 60°C$) and then stirred to promote syneresis (fluid moving from the curds to the whey) (Papademas & Bintsis, 2017, p. 133). When the cheesemaker (subjectively) judges that the cooking process is done, the curds and whey are separated, with the curds being pressed to form the cheese. (Papademas & Bintsis, 2017, p. 135). As a final production step the cheese is salted by submerging it into brine for the desired duration. This duration may vary for different cheeses, as salt uptake of the cheese depends on multiple variables, like its size and shape or the ratio of brine volume and cheese weight (Bintsis, 2007, p. 271). Now the cheese is ready to mature and be packaged for transportation.

Role of Curds in Cheese Production

The curds have a large impact on the quality of the eventual cheese that is made. An important aspect of the curds is their size. The size of the curds determines the amount and velocity of whey expulsion from the curds to the whey (Law & Tamime, 2010; Whitehead & Harkness, 1954). This happens not only during cutting, but also further continues during the stirring/cooking step (Renault et al., 1997).

Moisture Content Since the whey expulsion from the curds determine its moisture content, the size of the curds

after cutting have a big influence on the final moisture content of the cheese that is produced (Iezzi et al., 2012; K. A. Johnston et al., 1991). When the curds are cut to a relatively small size, they have a higher surface area, allowing for more whey expulsion (Hickey et al., 2015). Additionally, the whey in the curd has a smaller distance to travel to escape it, further allowing whey expulsion (McSweeney et al., 2017, p. 160). As a result, smaller curds shrink more rapidly than larger curds (Panthi et al., 2019). Based on these characteristics, cutting the curds to a smaller size yields a cheese that has a lower moisture content (harder cheese), while larger curds result in a higher moisture cheese (softer cheese) (McSweeney, 2007, p. 75). The moisture content has an influence on the required maturing time of the cheese, with a harder cheese needing more time to mature (McSweeney et al., 2017, p. 17). Additionally, the moisture content also has a big influence on the final cheese properties like texture, as it is the main contributor to the cheese hardness (McSweeney et al., 2017, p. 5).

pH value Almost all of the lactose from the milk, ~ 98%, is lost in the whey during cheese production (Bansal & Veena, 2022). The remaining lactose in the curds, however, is of big importance on the resulting cheese properties. Cutting the curds into smaller sizes results in more syneresis, which results in a lower lactose concentration in the curds. This lactose is turned into lactic acid by the starter lactic acid bacteria (Bansal & Veena, 2022). As a result of this lactic acid production, the pH value of the curds decrease (Shakeel-Ur-Rehman et al., 2004). Cheese pH value in turn has a big influence on its textural properties (Lawrence et al., 1987), but also influences flavour perception (Papademas & Bintsis, 2017, p. 47).

Curd Fines Since maturing time has a significant impact on flavour development (Muir & Hunter, 1992; Piggott & Mowat, 1991; ScienceLearningHub, 2012), the cutting procedure has a big effect on the desired cheese production. Additionally, the size of the curds at cutting not only have a substantial influence on the direct cheese characteristics, but also on the yield of the cheesemaker. Curds are referred to as curd fines, when they are small enough to get carried away by the whey and wash water (Cross et al., 1977). If the curds are cut too small, it may result in too many of such curd fines and thus a reduced cheese yield (Law & Tamime, 2010). On the contrary, if the curds are cut too large, they may shatter during the consequent stirring process (K. A. Johnston et al., 1991), as the curd particles are still very fragile when the stirring process starts (P. Fox et al., 2017, p. 319). This can also result in curd fines, reducing yield. Additionally, the fracture leads to a large surface area of the curd (in contrary to a clean cut), which may lead to high fat losses from the curd to the whey (K. A. Johnston et al., 1991) or cause graininess (Marth & Steele, 2001, p. 356). The higher fat loss can eventually have an influence on the taste of the cheese (McSweeney et al., 2017, p. 883, fig. 34.13).

Autonomous Cheese Production

For the reasons mentioned before, it is essential that the cutting procedure is observed during cheesemaking and that the curd size is monitored throughout. During manual cheesemaking, an expert decides the stopping point of cutting, often based on the size (and shape) of the curds (Aldalur et al., 2019). When the whole cheesemaking production process is done autonomously, the cheesemaker needs to invest less (or no) time in the cheese production process, allowing them to invest more time in other aspects of the farm. However, this also takes away the expert eye of the cheesemaker which is required to keep an eye on the production process and the final cheese quality. Currently, the cutting time in the Lely cheese factory is set at a predefined value. However, being able to adjust the cutting time based on the curd size is essential to create consistent cheeses and gain better control over the produced cheeses. This means that there needs to be an autonomous method to replace the manual decision making. This research therefore focuses on an autonomous method to decide the stopping time of the cutting process during autonomous cheese production. More specifically, the following research question will be investigated: *is it possible to control curd sizes during an autonomous cutting process using image acquisition in combination with AI techniques and, if so, what are the visible effects of varying the curd sizes on the cheese pH value, moisture content and yield at the Lely cheese factory?*

Vision in Cheesemaking

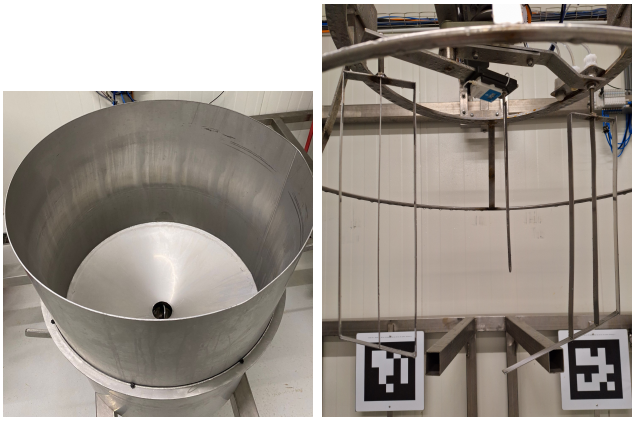
Currently, in the cheese industry there is a shift to automation and live monitoring of the production process, in order to improve the quality of the cheese (Fagan et al., 2007). Some methods of computer vision have already been used in cheesemaking. Mateo et al. (2010) have used traditional image segmentation methods to determine the syneresis of curds. They used a camera that was mounted to the vat to capture images of the curds and whey during the stirring process. Their segmentation method consisted of gray scaling the image, followed by deblurring and contrast enhancement. Finally Mateo et al. (2010) used four different methods, namely threshold, first order grey level statistics, grey level co-occurrence matrices and fractal dimension. Everard et al. (2007) have used a similar method, along with the R, G and B values of the curd/whey mixture to determine syneresis. However, using computer vision to control curd sizes autonomously during cutting has not been investigated before.

II. METHOD

In this section, the used materials and methods will be explained. First, the experimental environment will be discussed, followed by the curd detection algorithm. Finally, the experiment layout will be discussed.

Cheese Recipe

To ensure reproducibility, milk with a consistent composition is necessary. Milk composition can be cow dependent, by e.g. genetic factors or stage of lactation (P. F. Fox et al., 1998, p. 246). To minimise variation in milk, pasteurized milk from a farm where the milk is mixed together will be used. The milk will enter the cutting procedure at a constant temperature of 30 °C and a fixed quantity of 110 liters will be used for every cheese. The temperature of the room is kept at a constant 17 °C. As starter culture, 28 gr of Ceska®Star starter culture (DSM, 2023) is added to the milk, per cheese. Additionally, $CaCl_2$, $NaNO_3$ and rennet is added to the milk mixture. This mixture is allowed to coagulate for 30 minutes, after which the cutting process starts. To cut the coagulum, a barrel and knife as shown in figure 1 is used. The height of the barrel is 750 mm with a radius of 300 mm. During the mixing process, 32 liters of whey are sucked out of the mixture and 14 liters of wash water is added. This draining process is performed twice. Pressing the curds is done for 105 minutes, at a maximum of 5 bar. The cheese is then allowed to rest for 1 hour, before entering the brining phase for 3 days.



(a) Barrel *(b)* Knives

Fig. 1

Image acquisition

The camera that is used for image acquisition is a $MER2$–630–$18GC$ from $GeT\ Cameras$ (get-cameras, 2023). The camera settings are set up such that the gain, white-balance and exposure are adjusted automatically, with an expected gray value of 90. To ensure robust and continuous image acquisition, the width and height of the images are down-scaled to 1920 by 1080 pixels. The frame rate of the acquisition is set to 5 FPS. The camera is positioned such that the curds and part of the cutting vat are visible, see figure 2. The barrel is partially visible in the lower right corner. To ensure sufficient lighting, a LED strip is added to the cutting process, which is turned on when a barrel is placed around the knives and turned off at the end of the cutting process. Figure 3 shows a schematic view of this set-up.

Fig. 2: Image Aquisition View

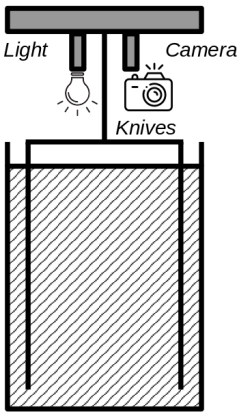


Fig. 3: Set-up Schematic

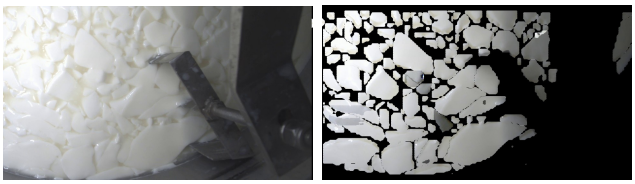
Curd Detection

To detect curds in the cutting vat, a YOLOv8 instance segmentation model is used. The chosen model is $YOLOv8x–seg$, which is the largest variant (Ultralytics, 2023). To perform curd detection, this model is trained on 16 images containing 2773 instances. These images are selected such that a wide variety of data is included in the training samples (e.g. knife visibility, shadow positions and a variety of curd sizes). Training was done on an NVIDIA RTX 6000 ADA GENERATION GPU for 20000 epochs, with a batch size of 1 and image size of 640. This model is trained to recognize three classes: curd, knife and barrel. For more detailed hyperparameters, see section VI-A.

Two main variables are extracted from the detections, namely the median size of the visible curds (in pixels) and the total number of curds that are visible within the frame. See figure 4 for an example of a frame with these two variables visible. The knives are visible within the camera's view and are, at certain frames, blocking the visibility of the curds, see figure 5. Therefore, the detected number of curds may, at some time points, have high variability.

(a) Acquired Image *(b)* Detected Curds With Median Curd Size and Number of Curds

Fig. 4



(a) Acquired Image With Knives In View *(b)* Resulting Curd Detections

Fig. 5

To compensate for this occurrence, an extrapolation of the number of visible curds is performed, based on the percentage of pixels that are occupied by a knife. See equation A.1 for this extrapolation. Additionally, a moving median filter with a window size of 25 is used to minimize noise created by the model inference inconsistencies. Inference is performed once per second, to ensure enough time for inference and calculations at each frame.

Curd Size and Number of Curds

The curd size in pixels is extracted by retrieving the instances and counting the number of pixels making up this instance. Since the general shape of the curds towards the ending phase of the cutting process can best be approximated as a circle, the number of pixels making up the instance are projected into a circular shape. The diameter of this circle will hereafter be defined as the curd size. With the assumption that the area of a pixel is known (in $\frac{mm^2}{pixel}$), then this projection into a circular shape can be made. See equation A.2, where d_{curd} is the diameter of the curd in mm. Additionally, the number of curds, represented as the total number of instances retrieved from inference, is also taken into consideration. This variable is helpful to gain more insight in the cutting process within the cutting vat (e.g. curd cutting rate and curd ripeness).

To assess the obtained results from the models curd predictions, the RMSE, precision/recall and balanced accuracy of the results are used. The RMSE is used to compare the error of the median filtered number of curds detected, with and without a correction based on the knife visibility. The error is defined as the difference between results from the algorithm and results obtained from manual image annotations (the ground truth). The RMSE is defined as shown in equation 1, where n is total number of samples and e is the error.

$$\text{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} e_i^2} \tag{1}$$

The precision and recall metrics can be used to give more in-depth insight into the predictions of a single class. These variables are defined as shown in equation 2, where TP is the number of true positives, FP the number of false positives and TN and FN the number of true and false negatives respectively (Padilla et al., 2020).

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

The curds and whey mixture dataset is not balanced (i.e. the class occurrences are not evenly distributed). An example can be seen in figure 6, where curds occupy 71% of all pixels.

Fig. 6: Curds occupy 71% of all pixels

If the model were to predict that all pixels belong to a curd, it would mean that its accuracy would equal 71%, giving a distorted indication of its performance. Therefore, the balanced accuracy measure is used, which can deal with unbalanced datasets better. Its definition is shown in equation 3 (Grandini et al., 2020).

$$TPR = \frac{TP}{P} \qquad TNR = \frac{TN}{N} \tag{3a}$$

$$\text{Balanced Accuracy} = \frac{TPR+TNR}{2} \tag{3b}$$

Dimension Conversion

The detection of curds happens on a pixel level (e.g. a curd has the size of 100 pixels). While this metric can serve as a way to compare curd sizes between various cheeses, it is not an intuitive unit for the farmer to set up the desired cutting size of the curds. Therefore, a transformation from curd size in pixels to the metric system is needed. Since the amount of milk in the cutting vat is not constant (a margin of 7.5 liters is used) and Lely currently adjusts the amount of milk based on seasonal

changes (which can vary by multiple liters as well). Based on these variations, the height of the milk in the cutting vat will vary significantly, which changes the conversion ratio from pixels to metric units, based on the height of the milk in the vat, and thus its distance from the camera.

To account for this, the known constant size of the cutting vat is used. The camera frame is positioned in such a way that a side of the barrel is visible, see figure 4a. This image is fed to the YOLOv8 network to perform inference with the desired class being the cutting vat. This inference is done with 100 acquired images, before the cutting process begins. If a pixel was detected to be a pixel in at least 70% of the images, this pixel is decided to be part of the barrel and is therefore added to the barrel mask. An example of a resulting mask is shown in figure 7a. With this mask, the points on the inner radius of the barrel are collected, see figure 7b.

(a) Barrel Mask

(b) Barrel Mask With Points on Inner Radius

Fig. 7

On these points, a least square circle fit algorithm is used (Bullock, 2006) to find the radius of the cutting vat, in pixels, at fluid height. From this radius in pixels and the known radius of the cutting vat (300 mm), a conversion ratio from pixels to metric units can be calculated at the distance that the curds will be visible, see equation 4, where D_{mm} and D_{pixel} are the curd diameter in millimeters and number of pixels respectively. $C_{MmPerPixel}$ is the constant conversion ratio for the current cutting vat.

$$C_{MmPerPixel} = \frac{D_{mm}}{D_{pixel}} \tag{4}$$

Since each cutting vat is filled with a varying amount of milk, and thus has a varying fluid height, this conversion is necessary for every cheese individually.

Independent Variable

To investigate whether vision techniques can be used to improve control over cheese properties through the cutting process at the Lely cheese factory, multiple cheeses will be made with various curd sizes, using the aforementioned vision technique. The remainder of this section will describe the methods for this experiment. Table I shows the cheeses that will be made with the various curd sizes. These values are determined by the smallest and largest curd sizes that can be used within the Lely cheese factory. To ensure comparability between the various curd sizes, however, possible inconsistencies of the dimension conversion will be eliminated by using a constant value for the mm to pixel conversion. This is possible, since a constant amount of milk is used between all cheeses that are produced.

TABLE I: Quantity of Cheeses With Curd Size

Curd size [mm]	5.0	5.5	6.0	6.5	7.0	7.5	8.0
Quantity [-]	3	3	2	2	2	3	3

Dependent Variables

During the experiment, various metrics will be collected, to gain insight into the effect of the curd sizes that are determined by the model. Generally, the assessment of cheese is often done by expert judges. However, their personal preferences and their lack of ability to assess samples without taste fatigue makes this assessment subjective (Farkye & Fox, 1990). Therefore, to monitor and assess objectively, chemical and physical analyses of cheese is done (Farkye & Fox, 1990). In this experiment, the following dependent variables will be measured: the moisture content of the cheese, its pH value and the amount of curd fines that were lost during the washing procedure (loss of yield). In section I the importance of these variables was discussed.

Moisture content Measuring the moisture content of a cheese can be done by weighing a part of a cheese, drying it in an oven and determining the lost mass of the cheese part. O'Keeffe (1984), for example, investigated the influences on moisture content of cheese by drying 1 gr of the cheese in an oven of 105 $°C$ for 5 hours. For the sake of consistency and comparability, the same method as previously used by the Lely cheese factory will be used; 13 days after the cheese has left the brining phase, 7 gr of cheese is dried for 24 hours in an oven at 105 $°C$.

Cheese pH To measure the pH of the cheese, a testo 205 digital pH meter (Testo, 2023) is used. After 13 days of resting a measurement of the cheese pH will be done at its core.

Curd Fines Lost To measure the curd fines, two methods are used. For the first method, the whey that is drained from the curds and whey mixture is first passed through a curved sieve of 200 μm. Next the whey is passed through a filter of 250 μm, followed by a finer filter of 200 μm, see figure 8. For every cheese the total amount of curd fines caught in these sieve and filters is weighed. For the second method, only the 200 μm curved sieve and the 250 μm filter is used.

Whey Fat Analysis To finally gain further insight into the influence of curd size on the curds, whey analysis is performed using a LactoScope FTIR FTA-3.x product analyzer (PerkinElmer, 2023). Whey samples are collected during production of the cheeses, taken at both draining procedures during mixing and right before pressing of

(a) Curved Sieve (200 μm)

(b) Circular Filter (200 μm and 250 μm)

Fig. 8: Filters To Measure Curd Fines

the cheese. The collected whey samples are tested for concentration of fats.

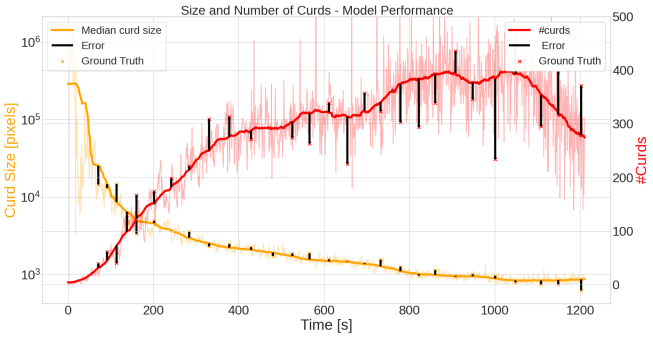
Realized Curd Sizes After Cutting To get a more quantifiable metric for the actual sizes of the curds, a sieving method can be used to quantify the realized curd sizes after cutting, as done by e.g. Kosikowski (1963) and K. Johnston et al. (1998). However, using such sieves to separate particles may cause large deviations (Igathinathane et al., 2009). Additionally, as these curds are still very fragile at this point in the process, it may cause breakage when done directly after cutting (K. A. Johnston et al., 1991) Therefore, a picture of the curds along with a tape measure is made, to get a sense of the real sizes of the cut curds.

III. RESULTS

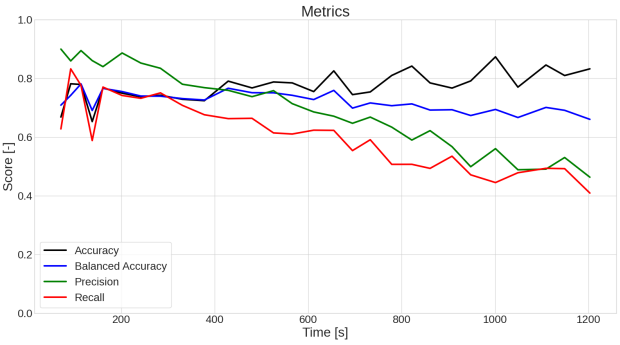
Performance

Model performance To assess model performance, various metrics will be used. First, to visualize the performance, manual annotations are made on the recording of the cutting program of a single cheese. The annotations are done in intervals of approximately 40 seconds. Based on these manual annotations, the ground truth of the median curd size in pixels and number of curds at various time points are determined. Next, the algorithm as described in section II is used to predict the same variables on the same recording. These results are shown in figure 9a without the number of curds correction and in figure 9b with the number of curds corrected for the knife visibility. The curd size is shown as the number of pixels a curd encompasses. Next, these predictions and manual annotations are used to calculate the metrics as discussed in section I. These metrics are shown in figure 9c.

(a) Model Performance



(b) Model Performance - Number of Curds Corrected



(c) Metrics

Fig. 9: Model Results

Noticeable from figures 9a and 9b is that the model can track the general trend of reduction of curd size and increase in number of curds. A first observation is the high variance in the number of curds detected in figure 9a, where the number of curds are not processed in any way. Figure 9b shows the effect of the correction of the number of curds, as described in section II. The variance decreases, with smaller peaks between close time points. It is however visible that there are some outliers where the number of curds are clearly overestimated at ~ 700 number of curds. Also noticeable from these figures is that with a correction on the number of curds, the number of curds seems to shift upwards. The largest number of curds that is passed through the median filter is at a maximum of ~ 360, while after the number of curds correction it

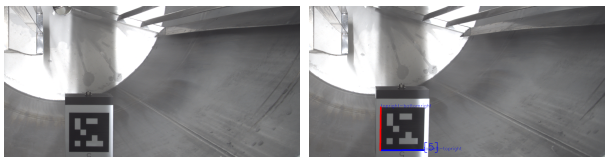
increases to ~400. The RMSE of the number of curds is found to be 56.92 pixels before the correction and 55.58 pixels after correction.

The curd size has a much smaller variance. The figure shows that, despite the number of curds increasing and the curd size decreasing over time, towards the end of the program, the number of curds that are detected start to decrease, while simultaneously the curd size starts to slightly increase towards the end of the program. Also noticeable is that the model start to overestimate the curd size towards the end of the cutting cycle.

Figures 9a and 9b also shows the error of the moving median filter of the model predictions compared to the manually annotated ground truth, in black vertical lines. The error of the number of curds shows, at certain samples, relatively high errors, with the errors also having a high variance between consecutive samples. On the contrary, the curd size error seems to be relatively stable in time. For the curd size, the mean error is determined by considering all the errors that are visualized with black bars in figures 9a and 9b. This mean error is found to be 282.2 pixels. However, also visible from figures 9a and 9b is that in the first 200 seconds, large errors are found. This is at a point that the curds are just being cut and a distinction between them is hard to make. If only the errors after 200 seconds are considered, the mean error becomes 74.2 pixels.

Figure 9c shows the performance metrics of the model. Initially noticeable is that the accuracy of the model seems stable over time, with fluctuations around 80%. However the balanced accuracy seems to show a slight downwards trend in time, with a final value of 65%, compared to an initial value of 75%. On the contrary, the precision and recall both show a very clear downwards trend in time, with the precision falling off from 90% to 46% and the recall decreasing from around 75% to 40%.

Dimension Conversion Performance To assess the accuracy of the used dimension conversion, first the dimension conversion is performed as discussed in section II. The distance from the top of the barrel to the height of the milk is measured. Next, to get a ground truth, a 5x5 dictionary ArUco marker of size 45 mm is placed at the same distance from the top of the barrel as the milk and its corners are detected using OpenCV. Based on the detected corners and the size of the ArUco marker, a ground truth pixel to mm conversion can be calculated. See figure 10 for the detections.



(a) ArUco Marker In Cutting Vat *(b)* ArUco Marker In Cutting Vat With Its Detections

Fig. 10

For this case, the model predictions resulted in a conversion value of $C_{pred} = 0.0269 \frac{mm^2}{pixel}$. The ground truth resulted in a conversion value of $C_{ground\ truth} = 0.0254 \frac{mm^2}{pixel}$, resulting in an error of $0.0015 \frac{mm^2}{pixel}$. Since the curd diameter scales by the square root of these conversion values, the error scales with 2.9 %, see equation A.3.

Combined Performance When taking both the curd detection and metric conversion into account, a measure for final performance can be produced. As discussed above, the mean curd detection error is found to be 74.2 pixels and the error for the metric conversion is found to be $0.0015 \frac{mm^2}{pixel}$. To interpret these errors in the context of curd sizes, a curd of diameter $d_{curd} = 6\ mm$ will be used as an example. When this curd is present in the acquired images, the algorithm may predict a curd size of 6.38 mm, see equation A.4.

By comparing the manually annotated data and the ground truth metric conversion to the algorithm predictions, an overview of the actual algorithm performance can be made. See figure 11 for these results.

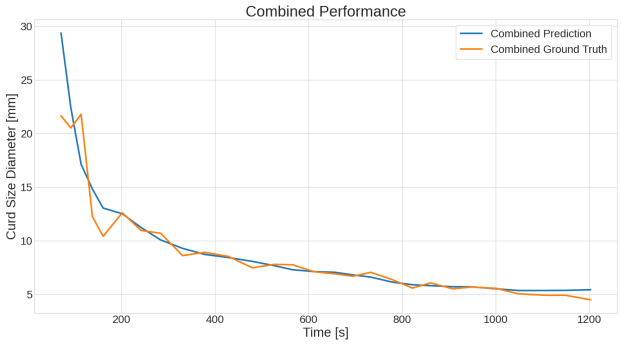
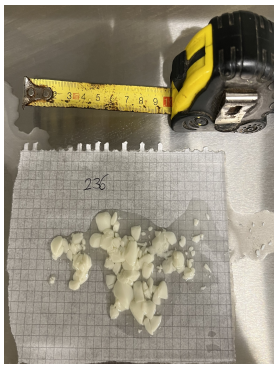


Fig. 11: Algorithm Performance

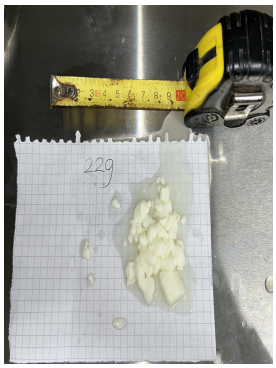
Here, the predicted curd sizes are obtained by combining both the YOLOv8 curd predictions with the predicted metric conversion value by circle fitting on the detected barrel. The ground truth curd sizes are obtained by combining the manually annotated curds with the ground truth metric conversion value. Visible from this figure is that the ground truth value makes some jittery movements at the start of the cutting program, up until ~200 seconds. The predicted curd size, however, seems much smoother at this same time frame, as the median filter smooths out jittery predictions, while at the same time still shows decreasing curd size. After the 200 seconds mark, the prediction and ground truth values are very close to each other, with the prediction often being in between the (slightly) varying ground truth values. After the ~1000 seconds mark, however, a divergence of both variables is visible. While the manually annotated curd sizes continue to decrease in size, the model detected curd sizes start to stagnate in size and, eventually, even a slight increase in predicted curd size is visible. This occurrence happens slightly above the 5 mm mark in curd size.

Realized Curd Sizes After Cutting Figure 12 shows the resulting curd sizes after cutting the cheeses at various

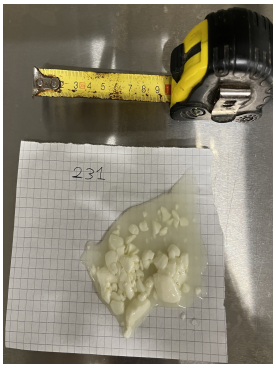
desired curd sizes. An example is shown for every curd size used. Visible from these results is that the curd sizes increase along an increase in desired curd size, confirming the results shown above. While the difference between two sequential sizes is harder to determine, see for example figure 12a and figure 12b, differences in size between curd sizes with multiple step sizes in between them are easier to distinguish. For example, the curds in figure 12a are significantly smaller than the curds in figure 12g. Another noticeable occurrence is that a large variance can be present within a size. In figure 12e, for example, there are some curds that are extremely large in size, but at the same time some are significantly smaller.



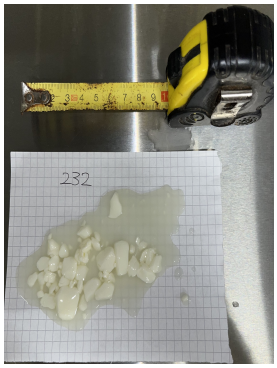
(a) 5.0 mm



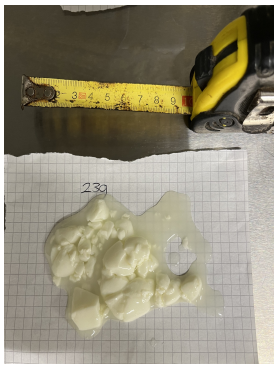
(b) 5.5 mm



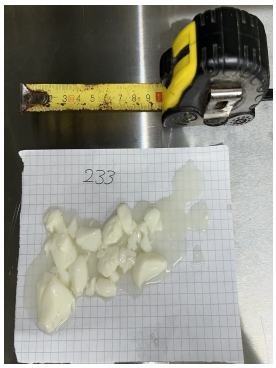
(c) 6.0 mm



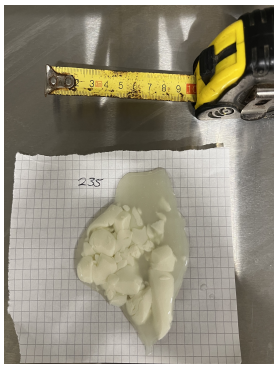
(d) 6.5 mm



(e) 7.0 mm



(f) 7.5 mm



(g) 8.0 mm

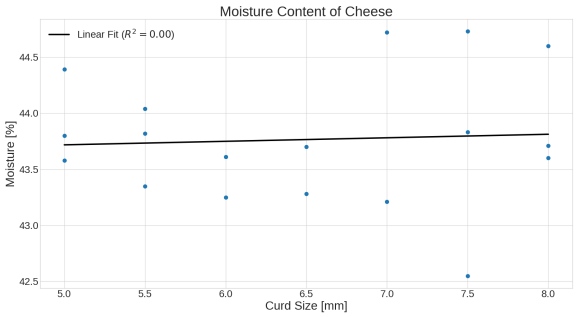
Fig. 12: Resulting Curds Based on Cutting Size

Cheese Properties Based on Vision Implementation

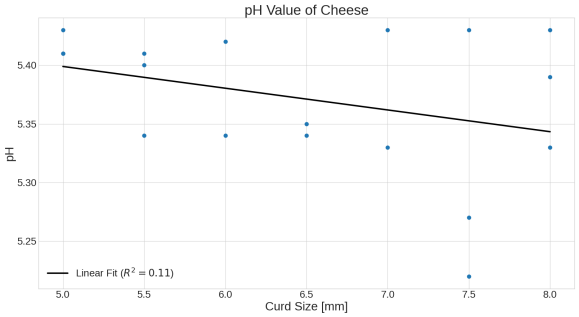
Figure 13 shows the measured cheese properties, based on the curd sizes that are determined by the algorithm in section II. Figure 13a shows the resulting cheese moisture contents. Visible from this figure is that the moisture content of the cheeses fluctuates around 43.5%. The moisture content, however, does not seem to follow a trend based on the curd size. While, at the initial increases in curd size, it seems that the moisture content slowly decreases, no eventual trend is found in the moisture content, with a coefficient of determination $R^2 = 0.00$.

Figure 13b shows the resulting cheese pH values, based on the curd size. Visible from this figure is that there, again, seems to be a slight inverse proportional relation between the curd size and the corresponding cheese pH value at the initial increase of curd sizes. Considering all curd sizes, however no linear trend can be found between the curd size and resulting cheese pH value, with a coefficient of determination $R^2 = 0.11$. Noticeable is the large spread in pH value at a curd size of 7.5 mm. Also visible from this figure is that at curd size of 8.0 mm, the pH values found were on the higher side, namely comparable to the pH values of curds of 5.0 mm.

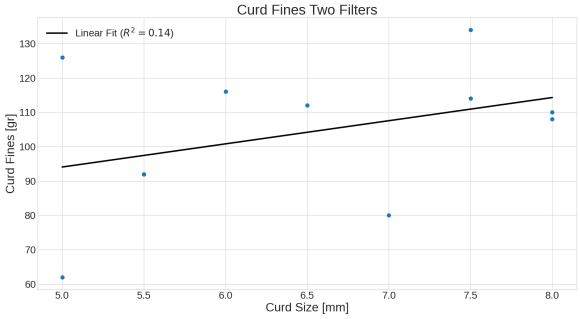
Figure 13c and figure 13d show the results of the amount of curd fines caught, using two ($R^2 = 0.14$) and three filters ($R^2 = 0.18$) respectively. Visible from comparing both figures is the difference in (slight) linear trends, with both a proportional and inverse proportional trend visible. Noticed during whey draining using all three filters, however, is that in the 200μm filter clogging occurred, resulting in drained whey overflowing and eventually not passing the filter and even washing away already caught curd fines. Only using two filters, without the 200μm filter, showed no signs of clogging and all whey was passed through the filters.



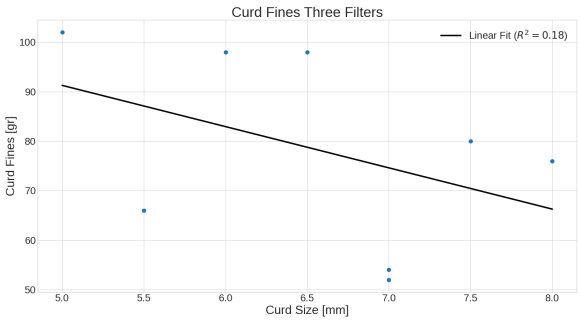
(a) Moisture Content of Cheeses



(b) pH Value of Cheeses



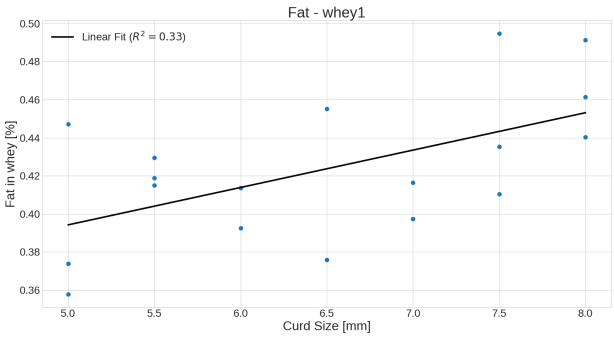
(c) Curd Fines With Two Filters



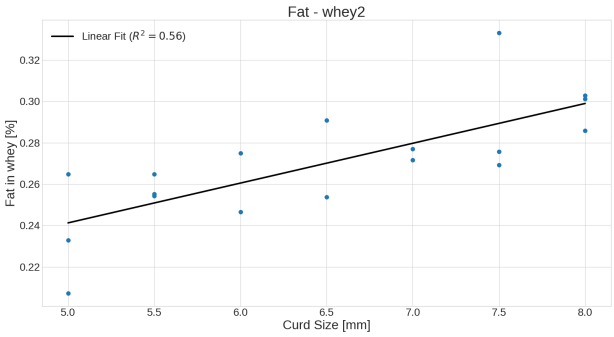
(d) Curd Fines With Three Filters

Fig. 13: Cheese Properties for Varying Curd Sizes

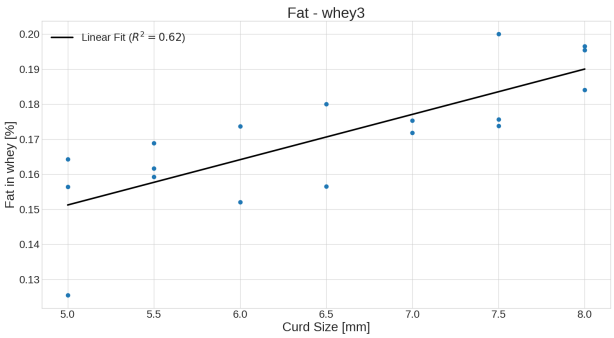
Fat Contents Whey Figure 14 shows the results of the measured fat concentrations by the LactoScope FTIR FTA-3.x product analyzer, depending on the desired curd size. *Whey 1* and *whey 2* are samples from the first and second draining of whey during the mixing process. *Whey 3* is the sample taken right before pressing of the cheese. Visible from all three images is that there seems a proportional relation between the curd size and fat contents in the whey. Comparing this trend between the three whey samples, the proportional relation seems to increase with the whey samples. This is also visible between figure 14a and figure 14b, as the spread within a curd size seems to decrease. Additionally, the coefficient of determination increases from whey 1 ($R^2 = 0.32$) until whey 3 ($R^2 = 0.63$), indicating that a stronger correlation forms between curd size and fat content in the whey. The absolute fat percentage within the whey is decreasing with each whey sample, from a maximum of $\sim 0.5\%$ in whey 1, to a maximum of $\sim 0.20\%$ in whey 3.



(a) Fat Whey 1



(b) Fat Whey 2



(c) Fat Whey 3

Fig. 14: Fat Concentrations in Whey

IV. DISCUSSION

Algorithm Performance

Curd Detection As seen in section III, the curd detection model performs sufficiently to detect curds in the cutting vat. From the model performance metrics, it was noticed that the balanced accuracy showed a slight decrease over time and the precision and recall curves showed a stronger decrease. Upon further investigation, it was found that the balanced accuracy slightly decreases over time due to the difference between all the positives and the true positives, along with the difference between all the negatives and the true negatives slightly diverging. As a result, the TNR increased, but at the same time the TPR decreased faster, resulting in the downwards trend of the balanced accuracy. Regarding the precision metric, it was found that the true positive value decreased over time, with the false positive value staying fairly constant. As a result, the decreasing trend in precision is visible. For the recall, a comparable occurrence happens, where the number of true positives decreases, while the number of false negatives stays relatively constant in time, hence also the decreasing model recall. These trends are caused by a lowered curd detection quality, which is expected to be caused by the phenomenon of sinking curds, which complicates the curds detection, even for human labelers.

Yet looking at the result in the context of curd detection, it can adequately track the general trends within the cutting vat, like decreasing curd size and increasing number of curds with time. In the first part of the cutting program (until the 100 seconds mark), however, the detections seem inconsistent, leading to jittery curd size predictions. The median filter does correct these predictions with large deviations, which still leads to a relatively smooth prediction of curd size at the beginning of the cutting program. The large variance in number of curds that are visible are to be attributed to the cutting knives obstructing the view within the cutting vat. Both without and with knife correction, however, comparable trends at similar time points are observed regarding the number of curds.

Correcting the number of curds based on the presence of knives within the image helps to reduce this aforementioned variance. Since the knife correction prevents the downward peaks due to the knife blocking the view of the curds, the median filter shifts upwards compared to the predictions without the knife correction. While the reduction in RMSE with respect to the median filter is moderate (from 56.92 to 55.58 pixels), the individual predictions show much lower variance. Some outlier values are present, where the number of curds is grossly overestimated. These values are expected to be due to the knife encompassing a large number of pixels in the image, without obstructing many of the curds. As a result, the extrapolation of the curds increase the already number of detected curds to an unrealistic value. It should be noted however, that these outliers are filtered out by the moving median filter.

Based on this result, a trend is noticeable for the number of curds visible within the vat, providing insights into the cutting program. One such insight is the stagnating number of curds between 550 and 700 seconds, while the knives are still rotating (visible in figures 9a and 9b). This could suggest that the RPM of cutting has become too low, resulting in the curds moving in front of the knife, preventing further cutting (K. Johnston et al., 1998). At the same time, the curd size still decreases, but at a noticeably smaller rate, visible from the lower slope at this time frame. This lower slope strengthens the possibility that the curds are no longer being cut in this time frame, while the curds are possibly still reducing in size due to syneresis. A different insight in the cutting program that can be gained from this analysis is that towards the final phase of the cutting (at $\sim$ 1050 seconds), the number of curds start to strongly reduce, which is not directly expected. At the same time, the curd size starts to stagnate again. This result is to be attributed to the curds shrinking and thus becoming heavier, making them sink in the whey (van der Haven & Oosterhuis, 1999, p. 81). As a result, the image acquired contains fewer curds that float in the fluid, see figure 15. The slightly sunken curds fall outside of the detection threshold for the model, which results in these curds not being detected, while only the larger floating curds are detected.

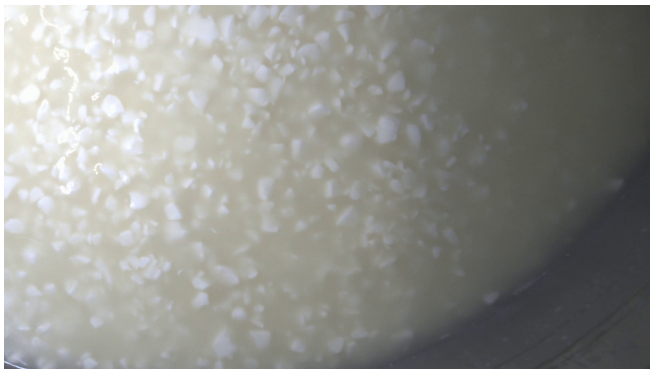


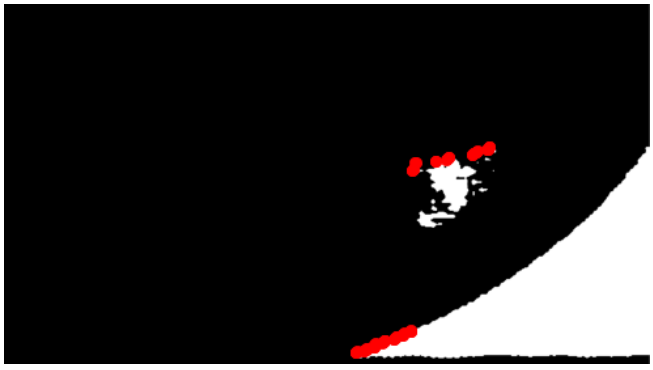
Fig. 15: Sinking Curds

Dimension Conversion As shown in section III, the dimension conversion algorithm produces a value with an error of 2.9% compared to the ground truth. While ideally an error of 0% is desired, this error value is within margins for the Lely cheese factory. The consistency of this dimension conversion, however, is not yet up to the desired standards, with $\sim$ 5% of the measurements not reaching a realistic $\frac{mm}{pixel}$ value. Mainly an incorrect barrel detection can result in such inconsistency, see figure 16 for an example of such an erroneous barrel detection. The curd/whey mixture is also being detected as part of the barrel, which hampers the inner radius detections, shown in figure 16c. These wrong inner points in turn lead to a wrong circle fitting radius and thus also an unrealistic

$\frac{mm}{pixel}$ value. While this error does not occur consistently, it does not yet allow for autonomous cheese production without human intervention. To prevent this error, further tuning of the model is recommended (e.g. better tuning of the hyperparameters), along with more barrel training images, possibly with additional shape fitting, to prevent the mask from taking unexpected shapes.

(a) Aqcuired Image

(b) Predicted Barrel Mask



(c) Predicted Inner Points For Circle Fit

Fig. 16: Erroneous Barrel Detections

It was also found that performing this barrel detection, while the knives are stationary, increased the chance of incorrect barrel masks. As discussed in section II, a pixel is only added to the barrel mask if this pixel was classified as a barrel in 70% of the detection attempts. If a stationary barrel is used, it increases the chances of having a erroneous prediction being repeated, as there is low image variability. Therefore, it is recommended for future use to perform this barrel detection only when the knives are non-stationary, e.g. during stirring of the starter culture.

Combined The combined performance of the curd size prediction and dimension conversion result in a sufficiently enough performance. The ground truth curd sizes and algorithm predicted curd sizes are very close during a cutting program. Noticeable is that the predicted values are in general in between the varying ground truth values. This confirms the usefulness of the smoothing using a moving median filter and the degree of accuracy of the algorithm results. Also here, however, it is noticeable that after a certain time point ($\sim 1000\ s$) the ground truth and prediction values start to diverge. This is again to be contributed to the sinking of the curds, yielding in less detectable curds and curds that are outside of the model confidence threshold. This poses a limitation for the minimum curd size that can be reached with this algorithm, as the curd ripening, and thus sinking, is not taken into account. Further tuning of the network (e.g. the confidence threshold value) may diminish this effect, possibly with a variable confidence threshold value at the end of the cutting process if a smaller curd size than 5 mm is required. It should be noted, however, that a curd size of $\sim 5\ mm$, which can still be tracked, is already smaller than the ideal curd size for the Lely cheese factory.

Realized Curd Sizes After Cutting It was found that the realized curds after cutting follow the trend of an increased size, with an increased desired curd size. It should be noted, however, that these samples are a small percentage of all the curds within a cutting vat, and thus may not be completely representative for the whole vat. Since the curd shapes are very irregular and the algorithm squashes their area into a circular shape, it is difficult to compare their final size with the desired ones.

A noticeable observation that was not expected, however, is the high variance in curd size within the samples. Upon further examination, it was found that during the cutting process, a possible dead point of the knives is detected near the midpoint of the cutting barrel, see figure 17. It is found that these curds gathered in the middle move with the rotating knives, while staying in front of them. This observation may explain the high variance within the samples after the cutting process, as not all curds are being cut equally. Additionally, this observation also has big implications on the dependent variables of this research, as this variance within a curd size prevents a homogeneous distribution of curd sizes within a cheese, while a homogeneous distribution is expected to control cheese properties. It is therefore strongly recommended for the Lely cheese factory to investigate the implementation of possible RPM changes on this high variance, as an increased RPM may prevent the curds from

floating away in front of the knives (K. Johnston et al., 1998).

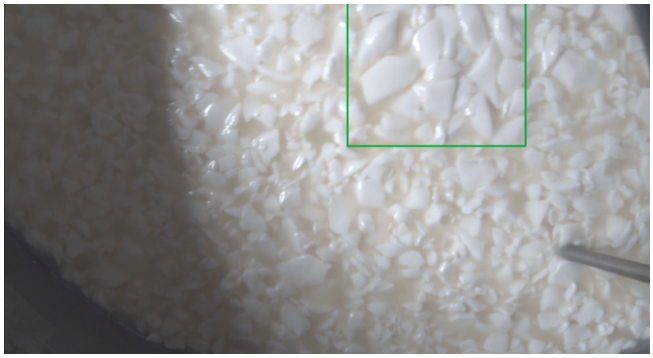


Fig. 17: Dead Point During Cutting Procedure

Cheese Properties Based on Vision Implementation

As discussed in section III, no recognizable trends are visible with cheeses made with different curd sizes at the Lely cheese factory. While, after the discovery of the dead point of the knives, the curd size distribution within a cutting cycle seems to have a larger than expected variance, still some trends would be expected with the created cheeses. The expected results would be, with increasing curd size, an increase in moisture content, a decrease in pH value and an increase in curd fines. The obtained results do not reflect this trend.

As for the curd fines, it is apparent that the results between two and three filters are not comparable. It was found that using three filters, however, resulted in regular clogging of the filter with a $200\mu m$ mesh and overflowing of the whey. This drainage resulted in not all whey being passed through the filters and thus do not reliably reflect all the fines within the drained whey. The increasing number of curds fines with using only the sieve and $250\mu m$ mesh, is better in line with expectations. It should be noted, however, that due to the split of data between two and three filters, there are too few data samples with only two filters to draw a strong conclusion. It is therefore recommended for the Lely cheese factory to only use the sieve and $250\mu m$ filter in the future.

For the moisture content of the cheeses, it would be expected to increase in value with an increased curd size. The findings, however, indicate at an initially lowering moisture content, with an increased spread at larger curd sizes ($> 6.5\ mm$).

Regarding the pH value of the cheeses, it is again initially decreasing with an increased curd size, with again an increased spread at larger curd sizes ($> 6.5\ mm$). This reoccurring trend in larger curd sizes showing a spread may indicate shattering of (some of) the curds at these larger sizes, resulting in this observed variance.

The fat content in the whey, however, does show a trend that is expected, that is that mainly larger curds may fracture during the mixing process and thus lead to a higher fat loss to the whey (K. A. Johnston et al., 1991).

A possible cause for the inconsistent and contradictory cheese properties may be attributed to a higher than usual mixing speed and the aggregate of small recipe variations in the production of the cheeses, which will be discussed next.

Due to a technical issue at the factory, the mixing speed was deviant during this study. While at the regular mixing speed a periodic oscillation of the mixing would be finished in 60 seconds, it was discovered that during this experiment, the period was, mistakenly, completed in a time of 6 seconds. This reduction in period duration results in a much higher RPM of the mixing pins, which in turn results in a considerably rougher mixing process. This notably rough mixing could possibly result in more curds shattering, which could negate the differences of the incoming curds after the cutting process.

It was found that there were, despite intensive efforts to ensure consistency, small recipe variations between the cheeses. One of such variations is the coagulation temperature. Figure 18 shows the coagulation temperature variations for the cheeses produced within this study. This temperature was measured manually right before the coagulation started. While the majority of the cheeses have a coagulation temperature of $30.3°C$, a large spread is found between temperatures of $30.4°C$ and $29.8°C$, with some additional outliers. This coagulation temperature has an effect on the coagulation rate and thus the firmness of coagulum at the time of cutting (Guinee et al., 1994; McSweeney et al., 2017, p. 836), consequently influencing the syneresis (Straatsma & Heijnekamp, 1988) and thus also cheese moisture content and pH value.

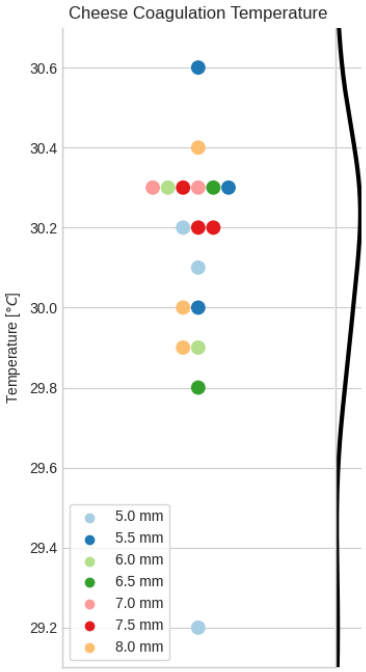


Fig. 18: Coagulation Temperature Variation

Furthermore, it was apparent during this study that it

was challenging to add the starter culture at the exact same time for every cheese produced. Between all cheeses, a variance of ~5 minutes was present in timing of starter culture addition. Moreover, the amount of starter culture was weighed at a scale with a resolution of 1 gr, with a desired amount of 28 gr. As the lactic acid bacteria in starter culture convert lactose into lactic acid (McSweeney et al., 2017, p. 871), with a higher dose and earlier addition of starter culture, more lactose is transformed into lactic acid, lowering the pH of the mixture. This could have contributed to the unexpected pH values of the created cheeses in this study. The role of starter culture, however, is more than just acidification, as it also has an effect on the biochemical and physical transformations during ripening (McSweeney et al., 2017, p. 644). While the extent of these effects on the cheeses produced in this study is not yet clear, it is noteworthy to mention for future cheeses produced at the Lely cheese factory.

Finally a form of variance between the different cheeses produced within this experiment is the duration between the cutting process and the mixing process. At the Lely cheese factory, logistical limitations posed it impossible for multiple cheeses to be at the mixing process at the same time. Therefore, if one cheese was finished with the cutting process while another cheese is still in the mixing process, progress of the former cheese is not possible. The curds and whey mixture at the cutting process, however, may not remain stationary, as this highly increases the chances of the curds clumping together, negatively influencing the further production of this cheese. The Lely cheese factory therefore uses a waiting procedure at the cutting process, in which the knives at the cutting station are rotated in counter clockwise direction, essentially mixing the curds and whey mixture. While this is a satisfactory method to ensure continuous autonomous production of cheese, it does hamper reproducibility at the curd level, as this additional mixing encourages syneresis and thus shrinks curds. Such a waiting process at the end of the cutting station has also occurred during the production of the cheeses within this study, which may have contributed to the inconclusive results in cheese properties.

V. RECOMMENDATIONS

Based on the obtained insights from this study, the following recommendations are made to further optimize the control over cheese properties and reproducibility within the Lely cheese factory. Firstly, while the detections of the curds are currently adequate, further increasing the training data for the YOLOv8 network is recommended, to get more consistent dimension conversions. This will also qualitatively improve the acquired data from the cutting process. Next, performing the barrel detections at a time where the knives are non-stationary is recommended, for improved dimension conversion consistency. If curd sizes smaller than ~5 mm are required, further research needs to be done into possibility of detecting these curds when ripening/sinking occurs. Regarding the measurements for the curd fines, it is recommended to not use the 200μm filter in the future, to get an accurate measurement of the curd fines.

It is further recommended to improve the consistency of cheese recipe to improve repeatability within the Lely cheese factory. Firstly it is recommended to control the milk temperature and thus minimize temperature spread when starting coagulation. Next, adding starter culture to the mixture at a more consistent timing is suggested, along with a higher resolution of control on the amount of starter culture added. Finally, further research into the prevention of logistics blocking cheese production is recommended, more specifically to avoid a cheese having to wait at the cutting station and thus already mixing the curds.

Finally, it is strongly recommended to investigate possibilities of reducing curd size variance after a cutting cycle. It is recommended to investigate the possibilities of diminishing the observed dead point within the cutting vat, e.g. by adapting to a different knife structure or using an increased RPM to prevent curds floating away in front of the knives.

VI. CONCLUSION

The goal of this research was to investigate an autonomous method to decide the stopping time of the cutting process during autonomous cheese production. To achieve this, a vision system is incorporated into the autonomous cheese factory of Lely. A YOLOv8 instance segmentation network is trained and evaluated. Next a conversion method from detected pixels to a comprehensible metric for the farmer is developed and evaluated. To finalize, an experiment is performed to investigate the practical implementation of this algorithm in the Lely cheese factory and its results are evaluated, with further recommendations for Lely.

It was found that the combination of curd detection and metric conversion could accurately follow the trend of the ground truth within a cutting schedule, until a curd diameter of 5 mm.

It was observed that the realized curds had a large spread in sizes and that cheeses that were produced with different curd sizes had conflicting results, where no recognizable trends in moisture content, pH value or amount of curds fines were found. Fat content in the whey, however, resulted in values more in line with the expectations, with a mixture of a larger curd size losing more fats to the whey, possibly due to more shattering. These inconclusive results are likely caused by inhomogeneous cutting of the curds and inconsistencies in the sensitive cheese making process. Recommendations are made towards Lely for direction of further research to improve repeatability in the cheese making process, in order to improve control over cheese properties, for cheeses made in the Lely cheese factory.

This research has shown promising results in autonomous curd size control. While further research needs

to be done within the Lely cheese factory to gain more control over the final properties of the produced cheeses, this study has set the foundation to autonomously control the curd size of the cutting process for autonomous cheese production.

References

Aldalur, A., Bustamante, M. Á., & Barron, L. J. R. (2019). Characterization of curd grain size and shape by 2-dimensional image analysis during the cheese-making process in artisanal sheep dairies. *Journal of Dairy Science*, *102*(2), 1083–1095. https://doi.org/https://doi.org/10.3168/jds.2018-15177

Bansal, V., & Veena, N. (2022). Understanding the role of ph in cheese manufacturing: General aspects of cheese quality and safety. *Journal of Food Science and Technology*, 1–11.

Bintsis, T. (2007, November). Quality of the brine. https://doi.org/10.1002/9780470995860.ch9

Bullock, R. (2006). Least-squares circle fit. *Developmental testbed center*, *3*.

Cross, S., Henderson, J., & Dunkley, W. (1977). Losses and recovery of curd fines in cottage cheese manufacture. *Journal of Dairy Science*, *60*(11), 1820–1823. https://doi.org/https://doi.org/10.3168/jds.S0022-0302(77)84107-6

DSM. (2023). *Ceska®star*. Retrieved November 6, 2023, from https://www.dsm.com/food-beverage/en_US/ingredients/dairy/cheese/ceska-star.html

Everard, C., O'Callaghan, D., Fagan, C., O'Donnell, C., Castillo, M., & Payne, F. (2007). Computer vision and color measurement techniques for inline monitoring of cheese curd syneresis. *Journal of Dairy Science*, *90*(7), 3162–3170. https://doi.org/https://doi.org/10.3168/jds.2006-872

Fagan, C. C., Castillo, M., Payne, F. A., O'Donnell, C. P., Leedy, M., & O'Callaghan, D. J. (2007). Novel online sensor technology for continuous monitoring of milk coagulation and whey separation in cheesemaking [PMID: 17854151]. *Journal of Agricultural and Food Chemistry*, *55*(22), 8836–8844. https://doi.org/10.1021/jf070807b

Farkye, N. Y. (2004). Cheese technology. *International Journal of Dairy Technology*, *57*(2-3), 91–98.

Farkye, N. Y., & Fox, P. F. (1990). Objective indices of cheese ripening. *Trends in Food Science & Technology*, *1*, 37–40. https://doi.org/https://doi.org/10.1016/0924-2244(90)90028-W

Fox, P., Guinee, T., Cogan, T., & McSweeney, P. (2017, January). *Fundamentals of cheese science*. https://doi.org/10.1007/978-1-4899-7681-9

Fox, P. F., Mcsweeney, P. L., & Paul, L. (1998). Dairy chemistry and biochemistry.

get-cameras. (2023). *Mer2-630-18gc*. Retrieved November 5, 2023, from https://www.get-cameras.com/6.3MP-GigE-Machine-Vision-Camera-Color-Sony-IMX178-MER2-630-18GC

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: An overview. *arXiv preprint arXiv:2008.05756*.

Guinee, T. P., Pudja, P. D., & Mulholland, E. O. (1994). Effect of milk protein standardization, by ultrafiltration, on the manufacture, composition and maturation of cheddar cheese. *Journal of Dairy Research*, *61*(1), 117–131.

Hickey, C., Auty, M., Wilkinson, M., & Sheehan, J. (2015). The influence of cheese manufacture parameters on cheese microstructure, microbial localisation and their interactions during ripening: A review. *Trends in Food Science & Technology*, *41*(2), 135–148. https://doi.org/https://doi.org/10.1016/j.tifs.2014.10.006

Iezzi, R., Locci, F., Ghiglietti, R., Belingheri, C., Francolino, S., & Mucchetti, G. (2012). Parmigiano reggiano and grana padano cheese curd grains size and distribution by image analysis. *LWT*, *47*(2), 380–385. https://doi.org/https://doi.org/10.1016/j.lwt.2012.01.035

Igathinathane, C., Pordesimo, L., Columbus, E., Batchelor, W., & Sokhansanj, S. (2009). Sieveless particle size distribution analysis of particulate materials through computer vision. *Computers and Electronics in Agriculture*, *66*(2), 147–158. https://doi.org/https://doi.org/10.1016/j.compag.2009.01.005

Johnston, K., Luckman, M., Lilley, H., & Smale, B. (1998). Effect of various cutting and stirring conditions on curd particle size and losses of fat to the whey during cheddar cheese manufacture in ost vats. *International Dairy Journal*, *8*(4), 281–288. https://doi.org/https://doi.org/10.1016/S0958-6946(98)00050-8

Johnston, K. A., Dunlop, F. P., & Lawson, M. F. (1991). Effects of speed and duration of cutting in mechanized cheddar cheesemaking on curd particle size and yield. *Journal of Dairy Research*, *58*(3), 345–354. https://doi.org/10.1017/S0022029900029927

Kosikowski, F. (1963). Some distribution patterns of cottage cheese particles and conditions contributing to curd shattering. *Journal of Dairy Science*, *46*(5), 391–395. https://doi.org/https://doi.org/10.3168/jds.S0022-0302(63)89061-X

Law, B. A., & Tamime, A. Y. (2010). *Technology of cheesemaking (2nd ed.)* Blackwell. https://doi.org/https://doi.org/10.1002/9781444323740

Lawrence, R., Creamer, L., & Gilles, J. (1987). Texture development during cheese ripening. *Journal of Dairy Science*, *70*(8), 1748–1760. https://doi.org/https://doi.org/10.3168/jds.S0022-0302(87)80207-2

Marth, E., & Steele, J. (2001). *Applied dairy microbiology*. CRC Press.

Mateo, M. J., O'Callaghan, D. J., Gowen, A. A., & O'Donnell, C. P. (2010). Evaluation of a vat wall-mounted image capture system using image processing tech-

niques to monitor curd moisture during syneresis with temperature treatments. *Journal of Food Engineering*, *99*(3), 257–262. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2010.02.019

McClure, S. B., Magill, C., Podrug, E., Moore, A. M. T., Harper, T. K., Culleton, B. J., Kennett, D. J., & Freeman, K. H. (2018). Fatty acid specific δ13c values reveal earliest mediterranean cheese production 7,200 years ago. *PLoS ONE*, *13*(9). https://doi.org/https://doi.org/10.1371/journal.pone.0202807

McSweeney, P. (2007). *Cheese problems solved*. Elsevier Science.

McSweeney, P., Fox, P., Cotter, P., & Everett, D. (2017, September). *Cheese: Chemistry, physics and microbiology, volume 1: General aspects* (Fourth).

Muir, D. D., & Hunter, E. A. (1992). Sensory evaluation of cheddar cheese: The relation of sensory properties to perception of maturity. *International Journal of Dairy Technology*, *45*(1), 23–30.

O'Keeffe, A. M. (1984). Seasonal and lactational influences on moisture content of cheddar cheese. *Irish Journal of Food Science and Technology*, *8*(1), 27–37. Retrieved September 6, 2023, from http://www.jstor.org/stable/25558086

Padilla, R., Netto, S. L., & Da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. *2020 international conference on systems, signals and image processing (IWSSIP)*, 237–242.

Panthi, R. R., Kelly, A. L., McMahon, D. J., Dai, X., Vollmer, A. H., & Sheehan, J. J. (2019). Response surface methodology modeling of protein concentration, coagulum cut size, and set temperature on curd moisture loss kinetics during curd stirring. *Journal of Dairy Science*, *102*(6), 4989–5004. https://doi.org/https://doi.org/10.3168/jds.2018-15051

Papademas, P., & Bintsis, T. (2017, September). *Global cheesemaking technology: Cheese quality and characteristics*. https://doi.org/10.1002/9781119046165

PerkinElmer. (2023). *Lactoscope™ ft-a liquid dairy products analyzer*. https://www.perkinelmer.com/product/lactoscope-ft-a-liquid-dairy-products-analyzer-lafta

Piggott, J. R., & Mowat, R. G. (1991). Sensory aspects of maturation of cheddar cheese by descriptive analysis [Cited By :68]. *Journal of Sensory Studies*, *6*(1), 49–62. https://doi.org/10.1111/j.1745-459X.1991.tb00501.x

Renault, C., Gastaldi, E., Lagaude, A., Cuq, J., & De La Fuente, B. T. (1997). Mechanisms of syneresis in rennet curd without mechanical treatment. *Journal of Food Science*, *62*(5), 907–910. https://doi.org/https://doi.org/10.1111/j.1365-2621.1997.tb15004.x

ScienceLearningHub. (2012). *Creating different cheese characteristics*. Retrieved February 16, 2023, from https://www.sciencelearn.org.nz/resources/829-creating-different-cheese-characteristics

Shakeel-Ur-Rehman, Waldron, D., & Fox, P. F. (2004). Effect of modifying lactose concentration in cheese curd on proteolysis and in quality of cheddar cheese. *International Dairy Journal*, *14*(7), 591–597. https://doi.org/https://doi.org/10.1016/j.idairyj.2003.11.008

Silanikove, N., Leitner, G., & Merin, U. (2015). The interrelationships between lactose intolerance and the modern dairy industry: Global perspectives in evolutional and historical backgrounds. *Nutrients*, *7*(9), 7312–31. https://doi.org/https://doi.org/10.3390/nu7095340

Straatsma, J., & Heijnekamp, A. (1988). *Kwantitatieve invloed van diverse factoren op het kaasbereidingsproces*. NIZO.

Testo. (2023). *Testo 205 - one-hand ph/temperature measuring instrument*. https://www.testo.com/en-US/testo-205/p/0563-2051

Ultralytics. (2023). *Ultralytics yolov8 docs - segment*. Retrieved November 5, 2023, from https://docs.ultralytics.com/tasks/segment/

van der Haven, T., & Oosterhuis, H. (1999). *Rondom boerenkaas* (tech. rep.). Praktijkonderzoek Rundvee, Schapen en Paarden.

Whitehead, H. R., & Harkness, W. L. (1954). The influence of variations in cheese-making procedure on the expulsion of moisture from cheddar cheese curd [Copyright - Copyright Dairy Industry Association of Australia Jul-Sep 1954; Document feature - Tables; ; Last updated - 2014-05-18; CODEN - AJDTAZ]. *Australian Journal of Dairy Technology*, *9*(3), 103–107. https://www.proquest.com/scholarly-journals/influence-variations-cheese-making-procedure-on/docview/199492480/se-2

APPENDIX

A. Hyperparameters YOLOv8 model

The following hyperparameters are used to train the used yoloV8 instance segmentation network.

```
task: segment
mode: train
model: yolov8x-seg.pt
data: dataset.yaml
epochs: 20000
patience: 0
batch: 1
imgsz: 640
save: true
save_period: -1
cache: false
device: 1
workers: 8
exist_ok: false
pretrained: true
optimizer: auto
verbose: true
seed: 0
deterministic: true
single_cls: false
rect: false
cos_lr: false
close_mosaic: 0
resume: false
amp: true
fraction: 1.0
profile: false
overlap_mask: true
mask_ratio: 4
dropout: 0.0
val: true
split: val
save_json: false
save_hybrid: false
conf: null
iou: 0.7
max_det: 300
half: false
dnn: false
plots: true
source: null
show: false
save_txt: false
save_conf: false
save_crop: false
show_labels: true
show_conf: true
vid_stride: 1
line_width: null
visualize: false
augment: false
agnostic_nms: false
classes: null
retina_masks: false
boxes: true
format: torchscript
keras: false
optimize: false
int8: false
dynamic: false
simplify: false
opset: null
workspace: 4
nms: false
lr0: 0.01
lrf: 0.01
momentum: 0.937
weight_decay: 0.0005
warmup_epochs: 3.0
warmup_momentum: 0.8
warmup_bias_lr: 0.1
box: 7.5
cls: 0.5
dfl: 1.5
pose: 12.0
kobj: 1.0
label_smoothing: 0.0
nbs: 64
hsv_h: 0.015
hsv_s: 0.7
hsv_v: 0.4
degrees: 0.0
translate: 0.1
scale: 0.5
shear: 0.0
perspective: 0.0
flipud: 0.0
fliplr: 0.5
mosaic: 1.0
mixup: 0.0
copy_paste: 0.0
cfg: null
v5loader: false
tracker: botsort.yaml
```

B. Formulas

$$
\begin{aligned}
A_{knife} &= \sum pixel_{knife} \\
A_{knife,\ norm} &= \frac{A_{knife}}{image_{width} \cdot image_{height}}
\end{aligned}
\tag{A.1}
$$

$$
\begin{aligned}
A_{barrel} &= \sum pixel_{barrel} \\
A_{barrel,\ norm} &= \frac{A_{barrel}}{image_{width} \cdot image_{height}}
\end{aligned}
$$

$$
n_{curds,\ extrap} = \frac{n_{curds,detected}}{1 - \left(A_{knife,\ norm} + A_{barrel,\ norm}\right)} \cdot \left(1 - A_{barrel,\ norm}\right)
$$

$$
\begin{aligned}
A_{curd} &= \#pixels \cdot A_{pixel} \\
A_{curd} &= \frac{\pi d_{curd}^2}{4} \\
\pi d_{curd}^2 &= 4 A_{curd} \\
d_{curd}^2 &= \frac{4 A_{curd}}{\pi} \\
d_{curd} &= \sqrt{\frac{4 A_{curd}}{\pi}}
\end{aligned}
\tag{A.2}
$$

$$
\frac{\sqrt{C_{pred}} - \sqrt{C_{ground\ truth}}}{\sqrt{C_{ground\ truth}}} = 2.9\ \%
\tag{A.3}
$$

Assuming a real curd of 6 mm:

$$
\begin{aligned}
A_{curd,real} &= \frac{\pi \cdot d_{curd}^2}{4} = 28.28\ mm^2 \\
\#pixels_{curd,real} &= \frac{A_{curd}}{C_{ground\ truth}} = \frac{28.28}{0.0254} \approx 1113
\end{aligned}
\tag{A.4}
$$

Since the curd detection in pixels has a mean error of 74.2 pixels:

$$
\#pixels_{pred} = 1113 + 74.2 \approx 1188
$$

$$
A_{curd,\ pred} = \#pixels_{pred} \cdot C_{pred} = 1188 \cdot 0.0269 = 31.94\ mm^2
$$

Rewriting this predicted area to a circular curd diameter gives:

$$
d_{curd\ pred} = \sqrt{\frac{4 \cdot A_{curd\ pred}}{\pi}} = \sqrt{\frac{4 \cdot 31.94}{\pi}} = 6.38\ mm
$$

C. Literature Review

1 Introduction

1.1. Lely Cheese Robot

Lely is a firm that was founded in 1948. From its start, the company was invested in machines to help the farmer in agricultural and livestock farming related tasks. Nowadays it mainly focuses on robots to use in livestock farming. The company has created many robots, varying from simpler robots that push cattle feed into position for cows to eat, like the Lely Juno, to more complex machines like the Lely Astronaut that can autonomously milk cows.

Since 2017 The R&D department of Lely has been working on their next big robot to help the farmer; one that can autonomously make cheese, locally at a farm. This robot closely cooperates with the Lely Astronaut, using its fresh milk to directly create cheese at the farmer. The cheese project has already produced more than 150 cheeses during testing and is continuously improved until it is ready to present publicly. This research is aimed to make one more of such improvements on the cheese robot, to achieve its final goal of autonomously and continuously making cheese.

1.2. Cheese General

Cheese is consumed as a complementary part of a main dish, or even enjoyed on its own (Farkye, 2004). It is then not surprising that world production of cheese is more than 18 million tonnes per year and steadily increasing along with the cheese consumption (Fox et al., 2017, p. 8). Even more telling about its importance is that around 35% of all milk is turned into cheese (Fox et al., 2017, p. 7). Daily dairy intake has the supposed health benefits like a lowered risk of myocardial infraction for woman who eat a lot of cheese, a reducing chance of bone fractures from osteoporosis (Soustre & Rayr, 2012) or even have a reducing effect on the risk of hypertension (Huth et al., 2006).

1.3. Cheese History

The discovery of cheese started a long time ago, namely more than 7000 years ago (McClure et al., 2018). At that time it was customary to use animal stomachs to transport liquids like milk. When this happens, the (remainders of) rennet in the stomach lead to the coagulation of the milk, which made it possible to extract cheese curds and whey (Fox et al., 2017, p. 3), which is believed to have lead to the accidental discovery of cheese (Silanikove et al., 2015). Nowadays there are more than hundreds of cheese sorts being made (Fox et al., 2017, p. 2), based on this first discovery and the production processes have evolved during millennia.

1.4. Cheese Production

While all those types of cheese are produced slightly differently, the main procedure is the same. In general, the first step it to pasteurize the milk to deactivate pathogens. This can be done by heating the milk for a short time at $\sim 73°C$ or for a longer period at $\sim 64°C$. (McSweeney et al., 2017, p. 42). The next step is the acidification by adding starter culture bacteria to the milk, which produce lactic acid

from the present lactose to create the distinctive cheese flavour (McSweeney et al., 2017, p. 201). Next the rennet is added to the milk to coagulate it, forming it into a gel (McSweeney et al., 2017, p. 53). This gel is then cut in smaller pieces named curds. In doing so, the mixture now consist of pieces of curd (solids) and whey (liquid). When the curds are at the desired size, as subjectively determined by the cheesemaker, the mixture is cooked by heating (varying between $\sim 40°C$ and $\sim 60°C$) and stirred to promote syneresis (fluid moving from the curds to the whey) (Papademas & Bintsis, 2017, p. 133). When the cheesemaker (subjectively) judges that the cooking process is done, the curds and whey are separated, with the curds being pressed to form the cheese. (Papademas & Bintsis, 2017, p. 135). As a final production step the cheese is salted by submerging it into brine for the desired duration. This duration may vary for different cheeses, as salt uptake of the cheese depends on multiple variables, like its size and shape or the ratio of brine volume and cheese weight (Bintsis, 2007, p. 271). Now the cheese is ready to mature and be packaged for transportation.

1.5. Role of Curds in Cheese Production

The curds have a large impact on the quality of the eventual cheese that is made. An important aspect of the curds is their size. The size of the curds determines the amount and velocity of whey expulsion from the curds to the whey (Law & Tamime, 2010; Whitehead & Harkness, 1954). This happens not only during cutting, but also further continues during the stirring/cooking step (Renault et al., 1997). Since the whey expulsion from the curds determine its moisture content, the size of the curds after cutting have a big influence on the final moisture content of the cheese that is produced (Iezzi et al., 2012; Johnston et al., 1991). When the curds are cut to a relatively small size, they have a higher surface area, allowing more whey expulsion (Hickey et al., 2015). Additionally the whey in the curd has a smaller distance to travel to escape it, further allowing whey expulsion (McSweeney et al., 2017, p. 160). As a result, smaller curds shrink more rapidly than larger curds (Panthi et al., 2019). Based on these characteristics, cutting the curds to a smaller size yields a cheese that has a lower moisture content (harder cheese), while larger curds result in a higher moisture cheese (softer cheese) (McSweeney, 2007, p. 75). In general, the moisture content also decides the required maturing time of the cheese, with a harder cheese needing more time to mature (McSweeney et al., 2017, p. 17). ‘

Since maturing time has a significant impact on flavour development (Muir & Hunter, 1992; Piggott & Mowat, 1991; ScienceLearningHub, 2012), the cutting procedure has a big effect on the desired cheese production. Additionally, the size of the curds at cutting also have a big influence on the direct cheese characteristics, but also on the yield of the cheesemaker. When the curds are too small such that they get carried away by the whey and wash water, these curds are called curd fines (Cross et al., 1977). If the curds are cut too small, it may result in too many of such curd fines and thus a reduced cheese yield (Law & Tamime, 2010). On the contrary, if the curds are cut too large, they may shatter during the consequent stirring process (Johnston et al., 1991), as the curd particles are still very fragile when the stirring process starts (Fox et al., 2017, p. 319). This can also result in curd fines, reducing yield. Additionally, the fracture leads to a large surface area of the curd (in contrary to a clean cut), which may lead to high fat losses from the curd to the whey (Johnston et al., 1991) or cause graininess (Marth & Steele, 2001, p. 356). This is why it is essential that the cutting procedure is observed during cheesemaking and that the curd size is monitored throughout. During manual cheesemaking, an expert decides the stopping point of cutting, often based on the size (and shape) of the curds (Aldalur et al., 2019).

After the curds have formed a cheese, its quality needs to be assessed. While the opinion of experts does not always correlate with the opinion of consumers (McBride & Hall, 1979), it is still very important to be able to objectively ”measure” the quality of cheese. This is necessary to be able to determine the effect of varying the curd size in the cutting process and thus to compare the effects of different curd sizes on the different cheese sorts. This is however not straightforward, as perception and its communication vary between individuals (Fox et al., 2017). When eating a cheese, one can sense three dimensions, namely the character, quantity and acceptability. (Law & Tamime, 2010, p. 440). The character represents the flavour (e.g. salty). The quantity is the extent of presence of the character (e.g. strong). Finally the acceptability is whether a character is liked or not and to which degree (Law & Tamime, 2010). Along with the dimensions, one can also sense defects of the cheese. In general,

cheese grading is done based on defects of the cheese (McSweeney et al., 2017, p. 528). An example of such a grading form is shown in appendix A.

1.6. Autonomous Cheese Production

When the whole cheesemaking production process is done autonomously, the cheesemaker needs to invest less (or no) time in the cheese production process, allowing them to invest more time in other aspects of the farm. However, this also takes away the expert eye of the cheesemaker which is required during the production process, like deciding when to stop the cooking or when the curd size is sufficient and thus to stop the cutting process, like mentioned above, and to keep an eye on the final cheese quality. Currently the cutting time on the cheese robot is a predefined value. However, being able to adjust the cutting time based on the curd size is essential. This means that there needs to be an autonomous method to replace this decision making. Currently in the cheese industry there is a shift to automation and live monitoring of the production process, in order to improve the quality of the cheese (Fagan et al., 2007). This research focuses on an autonomous method to decide the stopping time of the cutting process by using vision in combination with AI techniques, to quantify curd size.

2

Vision in Cheesemaking

Some methods of vision have already been used in cheesemaking. Mateo et al. (2010) have used traditional image segmentation methods to determine the syneresis of curds. They used a camera that was mounted to the vat to capture images of the curds and whey during stirring. Their segmentation method consisted of gray scaling the image, followed by deblurring and contrast enhancement. Finally Mateo et al. (2010) used four different methods, namely threshold, first order grey level statistics, grey level co-occurrence matrices and fractal dimension. The simple threshold method was found to give a positive result. Everard et al. (2007) have used a similar method, along with the R, G and B values of the curd/whey mixture to determine syneresis. However, to the best of the authors knowledge, using computer vision to control curd sizes autonomously during cutting has not been investigated before.

To autonomously control curd size during cutting, it is essential to detect the curds in the cutting vat, in order to determine their size. Thereafter, the size can be compared to the desired curd size and the stopping time of the cutting process can be chosen accordingly. Vision techniques (i.e. camera above the cutting vat) can be used to acquire images of the curd and whey mixture. Next, a curd detection method is needed for this purpose. This chapter will dive into the options to detect curds and discuss their applicability.

2.1. Object Detection

To distinguish between the curds and whey, object detection can be used. This could be models such as YOLOv6 (Li et al., 2023) or Faster-RCNN (Ren et al., 2015). Object detection creates bounding boxes around objects. The downside, however, of this method is that the size of a curd is hard to determine from only a bounding box when the curd shape is irregular (which is possible in the case of curd/whey detection). Object segmentation can better approach the actual size of irregular shapes, see figure 2.1. It is visible that the bounding box around the irregularly shaped curd does not correctly approximate its shape. The segmentation however is much closer to the actual size of the curd, as it is not restricted in its shape. Therefore, segmentation techniques are better suited in this context and will be further investigated.

(a) Original image *(b)* Object detection *(c)* Object segmentation

Figure 2.1: Visualization of curd size difference between object detection and segmentation

2.2. Object Segmentation

Segmentation is a more accurate method to distinguish between the curds and whey. This is because, in contrary to object detection, with segmentation every pixel is labeled belonging to a class, resulting in a shape of the object (Lakshmanan et al., 2021). Image analysis using segmentation is being used in many different domains, like the medical (Cashman et al., 2002) or food industry (Valous & Sun, 2012). Two different types of segmentation can be differentiated, namely traditional and deep learning segmentation. In traditional segmentation, heuristics like thresholding with Otsu's method (Otsu, 1979) or the canny edge detection (Canny, 1986) are used to differentiate the various contents of an image. In deep learning segmentation, networks are trained using, for example, convolutional layers (Gan et al., 2021). Currently there is a big shift towards deep learning segmentation (Gan et al., 2021), with this approach often proving more powerful than the classical segmentation methods (Bandyk et al., 2021). Both traditional segmentation and segmentation by machine learning are considered for the purpose of curd/whey distinguishment.

2.2.1. Traditional Segmentation

Traditional segmentation methods separate parts of the image by performing mathematical operations on the image. These operations can range from comparing pixel values to a predefined value or even a graph search (Basavaprasad & Ravindra, 2014). The most popular methods of traditional segmentation techniques can be divided into two categories, namely region based segmentation and edge based segmentation (Marr & Hildreth, 1980; Rogowska, 2009). The edge based methods look for edges based on different characteristics between regions and are in general more noise sensitive, as they use local gradients. The region based methods look for homogeneity based on criteria to create regions and are in general more computationally expensive (Shaaban & Omar, 2009).

Edge based methods

For edge based segmentation, first a method of edge detection needs to be applied, followed by a linking strategy (Shaaban & Omar, 2009). To detect edges, a convolution can be used (Rogowska, 2009). Based on the filter (or kernel) that is used, various edges can be detected. Examples of filters that can be used are the Sobel filter or the Prewitt filter, see equation 2.1 for their horizontal segments (Chaple et al., 2015). One such method of edge detection is the canny edge detection (Canny, 1986), where an image undergoes Gaussian smoothing, the gradients in the image are found, non-maximum suppression is applied followed by double thresholding and hysteresis.

$$Sobel\ Filter_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix}, \quad Prewitt\ Filter_x = \begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix} \tag{2.1}$$

Another method of edge based segmentation is graph search. For this method, first an image gets mapped to a graph. There are multiple methods to do this, with the simplest map being a transformation of each pixel to a vertex in a graph (Morris et al., 1986). The weight of the edge connecting the pixels can be represented by the difference in their intensity, color, location or any other measure (Felzenszwalb & Huttenlocher, 2004). This way, the weight of the edges represent the similarity between the pixels (as a lower weight corresponds to a high similarity). The complete graph can be created by linking all vertices

together by edges, or by only linking neighbouring vertices (Morris et al., 1986). To get a segmentation, multiple methods can be used like a graph cut method or the shortest spanning tree method (Camilus & Govindan, 2012). With the shortest spanning tree method, for example, a shortest spanning tree of the graph is cut at its highest weights, creating partitions that differ the most from each other (Morris et al., 1986). The Felzenszwalb algorithm is an example of an efficient graph based algorithm where a graph is created from the input image, with edges, vertices and weights based on dissimilarity between pixels. New sub-graphs are being created such that these sub-graphs have a relatively small weight (thus are similar in pixel property) (Felzenszwalb & Huttenlocher, 2004).

Region based methods
Of the region based methods, one of the most popular method is the thresholding technique (which has multiple varieties, like the histogram thresholding (Celenk & de Haag, 1998) or Otsu thresholding (Otsu, 1979)), as it is simple and robust (Valous & Sun, 2012). The thresholding techniques can be divided into two categories, namely global thresholding and local thresholding. Global thresholding is the most popular method, where the same threshold is used along the whole image. For local thresholding the threshold value depends on the information of the surrounding pixels. Based on the threshold value and the pixel value, a binary segmented image can be created (Al-Amri, Kalyankar, et al., 2010).

A straightforward form of region based segmentation is a color based method. While this method is simple, it can still be useful for important segmentation applications. Chiu and Lin (2005), for example, have used color segmentation to perform lane detection, or Benallal and Meunier (2003) have used colors to segment road signs. With this method, the image can be separated into parts by using a lower and upper color threshold. Everything within this threshold gets included into the mask.

The K-means algorithm (Hartigan, Wong, et al., 1979) can also be used to separate regions in an image. This method uses K initial clusters. Next, all the points (in the context of an image these points are pixels) get assigned to their closest cluster. The center point of this cluster changes based on the average location of its points. These steps are repeated until convergence (Hartigan, Wong, et al., 1979). This method can also be used for segmentation using colors, namely by defining a cluster as a color value, assigning each pixel (which has a color value) to the K number of classes and performing the algorithm.

Another popular approach of the region based models is the region growing method (Adams & Bischof, 1994). This technique initially creates many small regions in an image (Chang & Li, 1994). These parts of an image are grouped together when their criteria are within a predefined value (Valous & Sun, 2012). This is done until there are no more merges remaining to form a homogeneous segment (Chang & Li, 1994). This method however does require manually entered seed points.

2.2.2. AI Segmentation

While there are positives of this traditional segmentation, like some of them being fast to compute or quick to tune (Bandyk et al., 2021), the downside is that from these methods it is hard to accurately apply segmentation if the shapes are complex (Bandyk et al., 2021). Deep learning can help to overcome this downside, as complex situations can be better handled (Wani et al., 2020). AI segmentation has improved massively over the years and has achieved the best performances on benchmarks (Minaee et al., 2022). Segmentation using AI can be divided into three sections, namely semantic segmentation, instance segmentation and a combination of both (panoptic segmentation) (Minaee et al., 2022). Semantic segmentation assigns each pixel to a certain class, while instance segmentation goes a step further and assigns each pixel to a certain unit of a class (like individual people) (Minaee et al., 2022). For this research, both semantic and instance segmentation are considered.

Many different models for segmentation have been proposed. Some of the popular state of the art models are Mask-RCNN, which is a model that makes detections on an image. After determining the objects in it, the model fits a mask within this detection, leading to segmentation (He et al., 2017) and DeepLabv3+, which is an improvement of DeepLabv3, where the former has an additional decoder module, to further improve the segmentation around object boundaries. (Chen et al., 2018). Another,

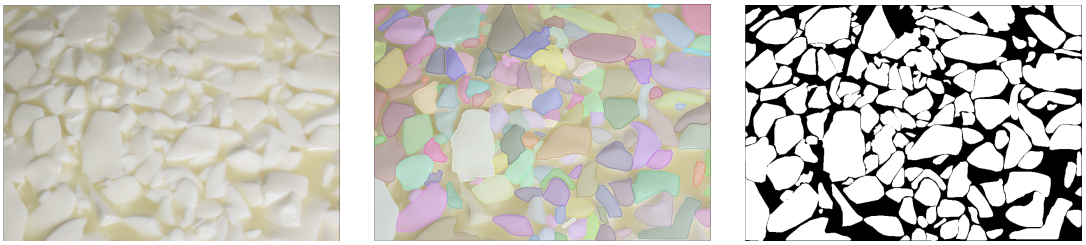
more recent, model is YOLOv8 (Ultralytics, 2023a). This is an improvement of earlier YOLO models, with performance and efficiency improvements along with new features (Ultralytics, 2023a). Some models that currently top the charts are InternImage-H (Wang, Dai, et al., 2022) or BEiT-3 (Paperswithcode, 2023; Wang, Bao, et al., 2022). To boost performance, often a backbone is used that is pre-trained on a dataset (Mo et al., 2022).

In addition to these state of the art models, a more simple machine learning algorithm can also be considered, like a random forest algorithm. The random forest algorithm is a combination of multiple trees, where each tree is generated using a randomly sampled vector. Each tree makes a prediction vote and the aggregate of these predictions decides the final decision made (Breiman, 2001). The random forest algorithm is an algorithm that can scale well with large volumes of information (Biau & Scornet, 2016). Additionally, it is an algorithm that is relatively fast to train and predict and has few tuning parameters (Cutler et al., 2012).

3

Segmentation Methods for Cheesemaking

In this chapter some of the previously discussed methods of segmentation will be applied on an image containing a mixture of curds and whey during the cutting procedure, to assess and compare the methods. First the classical segmentation methods will be shown followed by machine learning and AI methods. The runtime, along with its standard deviation, will be compared for each method. Additionally, a ground truth of the used image is made to help assess the results of the methods, see figure 3.1.



(a) Original Image *(b)* Overlay *(c)* Ground truth mask

Figure 3.1: Ground truth of the used image

Based on this ground truth, an accuracy, balanced accuracy, precision and recall measure is made to compare the methods. The accuracy metric is defined as the percentage of pixels that are assigned to the correct class (i.e. curd or whey). The precision and recall metrics are defined as shown in equation 3.1, where TP is the number of true positives, FP the number of false positives and TN and FN the number of true and false negatives respectively (Padilla et al., 2020).

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN} \tag{3.1}$$

It can be seen from figure 3.1c that this dataset is not balanced, as there are many more pixels belonging to the curd class compared to the whey class (70% vs 30%). If one of the methods were to predict that all the pixels belong to curd, it would still achieve an accuracy of 70%. Therefore, the balanced accuracy measure is used, which can deal with unbalanced datasets better. Its definition is shown in equation 3.2 (Grandini et al., 2020).

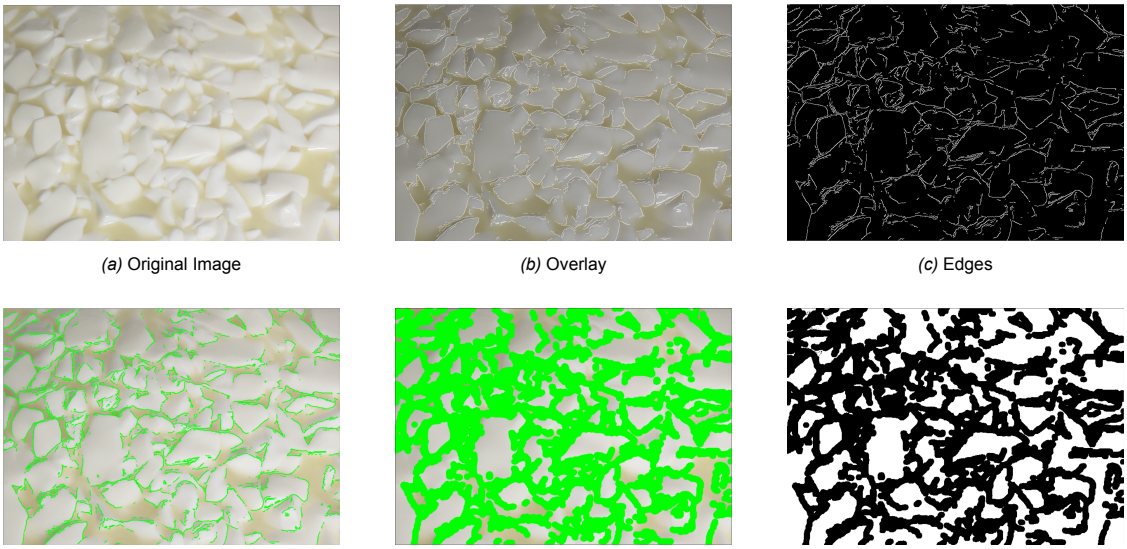
$$TPR = \frac{TP}{P} \qquad TNR = \frac{TN}{N} \tag{3.2a}$$

$$BalancedAccuracy = \frac{TPR + TNR}{2} \tag{3.2b}$$

After the methods and their metrics have been discussed, the final part of this chapter will consider the requirements of the cheese robot at Lely, to make a decision for the best suited segmentation method to use for autonomous control of the curd size.

3.1. Classical Segmentation

3.1.1. Canny Edge Detection



(a) Original Image *(b)* Overlay *(c)* Edges

(d) Contour of Edges *(e)* Dilated Contours *(f)* Mask

Figure 3.2: Canny Edge Detection to create a segmentation mask

Figure 3.2 shows the results of the canny edge detector on an image of curds and whey. For this method, the OpenCV function $Canny$ is used (OpenCV, 2023a). For this detection, the parameters that were used for the lower- and upper threshold were 25 and 30 respectively. These values were found by visual inspection of the resulting edges. It is visible from figure 3.2c that the edges are detected reasonably well, however, most edges are still open and no real distinction between curd and whey can yet be made. To get a mask to compare to the ground truth, the edges are dilated using OpenCV's $dilate()$ function (OpenCV, 2023b), with an elliptical 15 by 15 kernel, see figure 3.2e. Finally the mask is created by assigning the segments between the dilated contours as curds. This resulted in a balanced accuracy, precision and recall of 0.61, 0.81 and 0.51 respectively for this method. The average runtime of this algorithm is found to be 1.65 milliseconds, with a standard deviation of 0.129 milliseconds.

3.1.2. Felzenszwalb algorithm

(a) Original Image

(b) Felzenszwalb Segments

(c) Felzenszwalb Classes

(d) Mask

Figure 3.3: Felzenszwalb algorithm to create a segmentation mask

Figure 3.3 shows the result of the Felzenszwalb algorithm when used on a curd / whey mixture. For this application, the $scale$, $sigma$ and min_size parameters are set to 50, 0.5 and 50 respectively. It is visible that the edges of the curds are marked very accurately and that also good distinction between the curds are made. However, still an additional post-processing step is necessary to classify the individual curds. To achieve this, the average pixel color per segment is calculated. If the average color pixel in the segment is higher than $[0.77, 0.77, 0.75]$, the whole segment is assigned to curd and otherwise it is assigned to whey. This threshold is chosen by manually tuning, based on the visual assessment of the resulting mask. See figure 3.3d for the resulting mask of this method. This resulted in the following metrics: a balanced accuracy of 0.71, precision of 0.88 and recall of 0.61. The average runtime of this algorithm is found to be 582.77 milliseconds, with a standard deviation of 27.05 milliseconds.

3.1.3. Color based segmentation

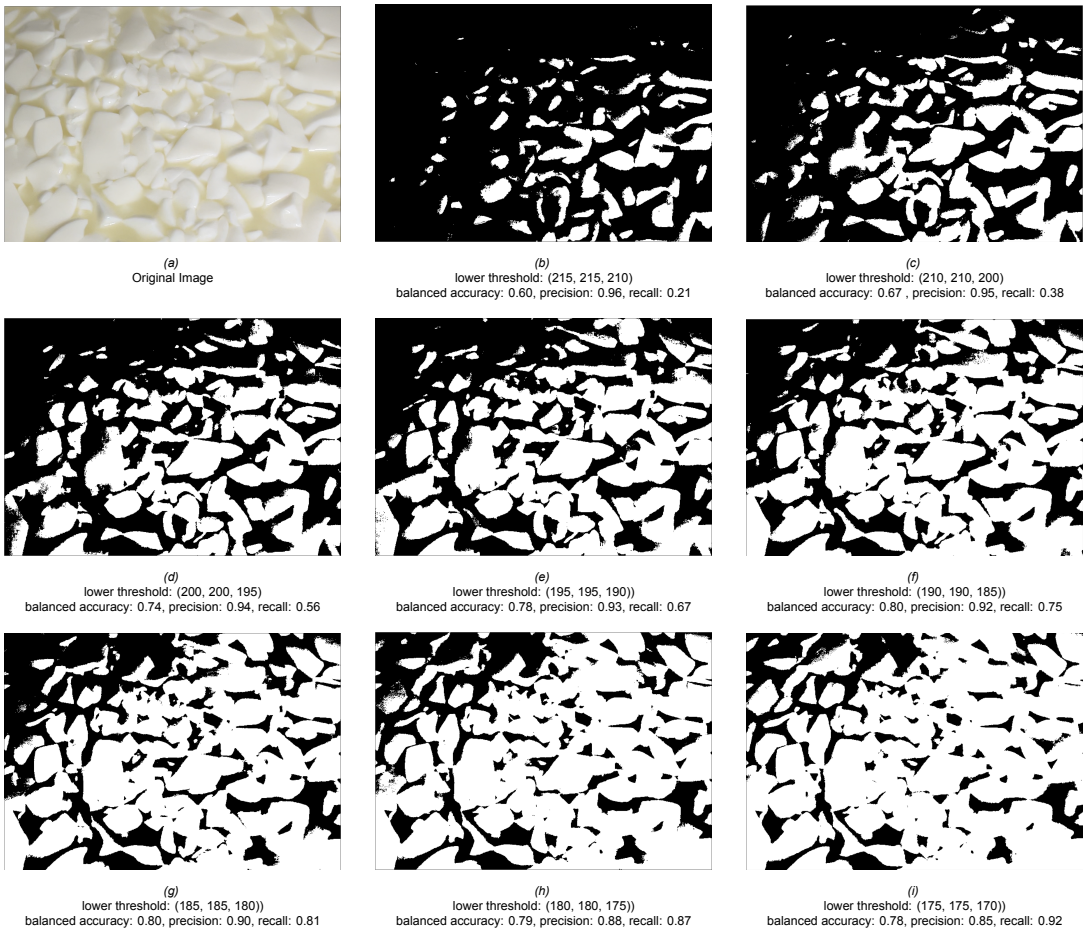


Figure 3.4: Color based segmentation for various lower threshold values

Figure 3.4 shows the application of color thresholding method on the curds and whey. This figure shows the results based on multiple lower color threshold values, while all of them use the same higher threshold value, namely $[255, 255, 255]$, as the curds are much closer to a white color compared to the whey. When a pixel's color value falls between the lower- and upper threshold, the pixel gets assigned to the curd class. It is visible that with the lower threshold getting closer to white (so [255, 255, 255]), less and less curds are detected. A threshold value of $[185, 185, 190]$ shows a good balance between too little curds detected and too much whey being included in the mask. Noticeable is that in the top left part of the image, where there is partly a shadow, the color intensity is different compared to the rest of the image, making it difficult to segment all the curds when there is variability of lighting conditions. Furthermore, an additional method of classification is needed to separate the individual curds (e.g. with the watershed algorithm (Beucher, 1979)). The average runtime of this algorithm is found to be 0.422 milliseconds, with a standard deviation of 0.093 milliseconds. For each threshold value, the accuracy, precision and recall are also measured and shown in figure 3.4. Noticeable is that with a lower threshold, the accuracy improves. This, however, seems to be caused by classifying almost all pixels as curds (since the ground truth image is dominated by pixels containing curds). This is also visible by the increasing recall and decreasing precision.

3.1.4. Otsu's thresholding

(a) Original Image *(b)* Mask

Figure 3.5: Otsu segmentation

Figure 3.5 shows the result of Ostu's segmentation method (Otsu, 1979) on a curds and whey mixture. After running the algorithm, a threshold value of 0.76 (normalized on a grayscale image) was returned. While at first sight it would seem that this method distinguishes correctly between curds and whey, upon closer look it is apparent that in the bottom right no real distinction between the two phases are being made. It is visible there that almost all whey in the bottom right corner is being segmented as curds instead of whey. This can be explained by the variation in lighting on the whole image, which means there is no one threshold value that can adequately segment all curds and whey using Otsu's method. The average runtime of this algorithm is found to be 4.74 milliseconds, with a standard deviation of 0.310 milliseconds. The balanced accuracy of this method was found to be 0.66, with a precision of 0.80 and recall of 0.75.

3.1.5. K-means algorithm

(a) Original Image *(b)* Mask

Figure 3.6: K-means segmentation for $k = 2$

Figure 3.6 shows the result of the K-means algorithm. For curds and whey, two groups can be distinguished, so $k = 2$ is chosen as the number of clusters in the image. Figure 3.6a shows that for $k = 2$ a similar result is shown as obtained with other classical segmentation methods, namely that the upper left corner with a shadow is not correctly segmented. Additionally the curds still need to be segmented into individual instances. The average runtime of this algorithm is found to be 423.5 milliseconds, with a standard deviation of 29.21 milliseconds. The balanced accuracy of this method was found to be 0.74, with a precision of 0.86 and recall of 0.77.

3.1.6. Region Growing

(a) Original image with seed point

(b) Ground truth

(c) Mask with threshold=1.78

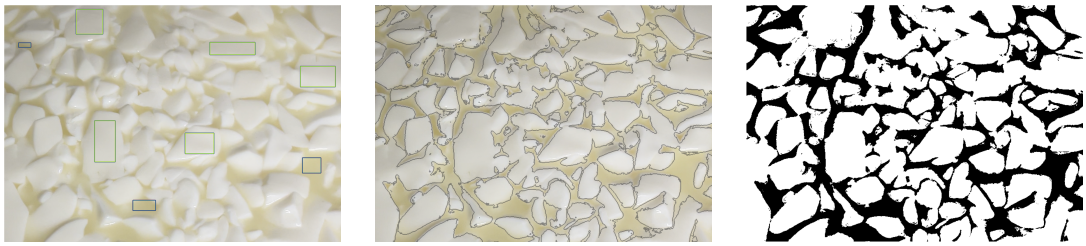
(d) Mask with threshold=2.90

Figure 3.7: Region growing segmentation

Figure 3.7 shows the result of the region growing algorithm on a curds and whey mixture. The results from this algorithm are dependent on the manually chosen seed point (marked with a red dot) and the chosen threshold tolerance. Both a relatively lower and higher threshold value are shown in figure 3.7. Noticeable is that for a lower tolerance, the segmentation stays near the seed point and fully encompasses the curd, although some not segmented (black) parts are visible within the curd and some small leakage to nearby curds is visible. With the higher tolerance value, however, significant leakage occurs, namely from the initial curd into other parts of image. The tuning of this parameter was done manually and would change based on the chosen seed point and the lighting conditions. By manually tuning this parameter, an optimal value of $tolerance = 1.78$ is found. The average runtime of this algorithm is found to be 3056.8 milliseconds, with a standard deviation of 155.2 milliseconds. To calculate the metrics for this method, only the ground truth of the curd containing the seed point is used, see figure 3.7b The balanced accuracy of this method was found to be 0.94, with a precision of 0.82 and recall of 0.88.

3.2. AI segmentation

3.2.1. Random Forest Segmentation



(a) Original image with labels

(b) Overlay

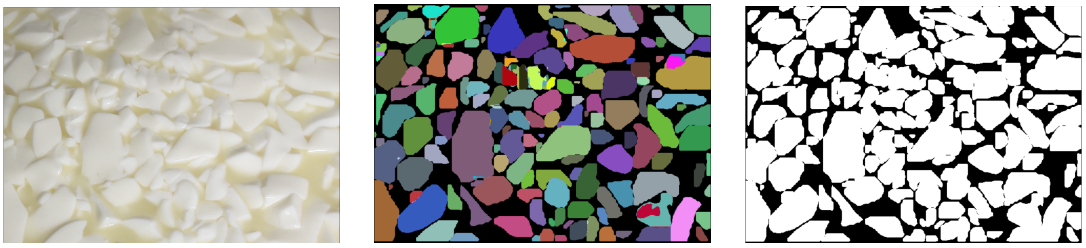
(c) Mask

Figure 3.8: Random Forest Segmentation

Figure 3.8 shows the result of the random forest algorithm on an image containing curds and whey. The original image along with the training labels supplied to the algorithm is depicted in Figure 3.8a.

The green labels represent the curds and the blue labels represent the whey. Figure 3.8c shows the result after training on these labels and predicting on the image. Noticeable is that, even though relatively few training samples were given, the result of the segmentation seems promising. Additionally, the shadow in the top-left corner of the image is handled much better, as the algorithm can make a distinction between curds and whey in that corner as well. However, it is visible that a large curd is being segmented as whey in that corner, so an improvement is still possible and required. The average training time, with the training labels as shown in figure 3.8a, is found to be 114.4 milliseconds, with a standard deviation of 11.24 milliseconds. The average prediction runtime of this algorithm after training is found to be 271.3 milliseconds, with a standard deviation of 5.23 milliseconds. The balanced accuracy of this method was found to be 0.82, with a precision of 0.88 and recall of 0.91.

3.2.2. YOLOV8 instance segmentation



(a) Original image *(b)* Instances *(c)* Mask

Figure 3.9: YOLOV8 segmentation

Figure 3.9 shows the result of instance segmentation, resulting from a prediction of a custom trained YOLOv8x-seg model (Ultralytics, 2023b). This model was trained on 6 images containing a mixture of curds and whey for 750 epochs, with a batch size of 6. A pre-trained model on the COCO dataset was used (Lin et al., 2014). The confidence threshold for inference was set at 0.25. Training of this network took 0.40 hours on an NVIDIA RTX 6000 Ada GPU.

This models not only results in a segmentation mask (as seen in figure 3.9c), but also distinguishes between the different curd instances, see figure 3.9b. Visible from the mask in figure 3.9c is that the model is capable of correctly differentiating between curds and whey in an image and that the edges between the instances are clearly defined. Also noticeable is that there is not a specific location within the image where performance is significantly worse, meaning that the model is able to correctly deal with shadows in the top-left corner. Figure 3.9b shows that the different instances of curds are being segmented from the input image. While this result looks promising, there is still performance to be found; it is noticeable that there are some overlapping colors, meaning that a single curd is being assigned to multiple instances. Based on the mask in figure 3.9c a balanced accuracy of 0.85, a precision of 0.90 and recall of 0.95 are found. The inference runtime of this algorithm was found to be 785.4 milliseconds, with a standard deviation of 42.11 milliseconds. This inference is run on the CPU to make comparisons with other methods possible.

4

Discussion

In this chapter the results seen in chapter 3 above will be discussed, to choose an ideal method to continue for curd-whey segmentation. First, the metrics will be looked at and compared, followed by a section that looks at the results from the perspective of the cheese robot at Lely, taking into account its requirements.

4.1. Metrics

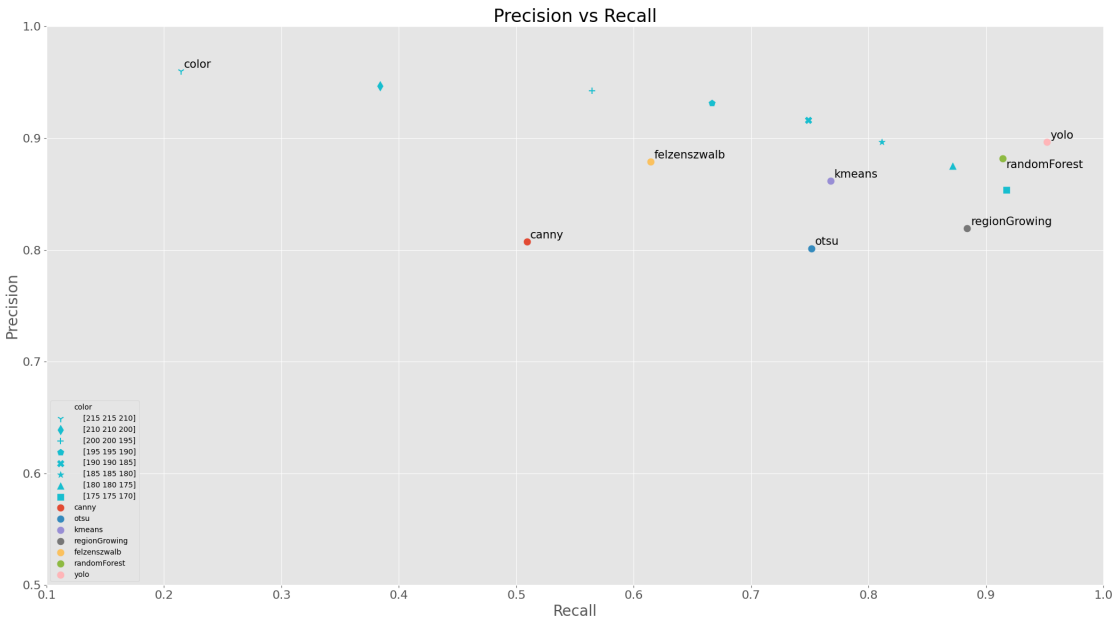


Figure 4.1: Precision vs Recall

Figure 4.2 shows the recall vs precision plot, as defined at the beginning of chapter 3, for the various methods. Ideally the precision and recall will both approach 1, meaning that the goal for each method is to reach the top right corner of the figure. Visible is that for color thresholding, at a high threshold value of $[215, 215, 210]$ (so a value closer to white), the method scores a high precision with a lower recall. This means that the method wrongly predicts many pixels to belong to whey and thus scores a low recall. By lowering this threshold value (e.g. to $[175, 175, 170]$) the inverse happens, as the recall increases but the precision decreases. This can be attributed to the fact that the color thresholding method now categorizes many pixels as curd instead of whey. This occurrence is also visible from the masks in figure 3.4. Also noticeable about the color thresholding is that even though the precision and

recall from some threshold values are relatively high, the curds in the corresponding masks are not easily distinguishable (e.g. figure 3.4i).

While the Canny edge detection scores a precision of 0.8, its recall is low to such a degree that, even with further tuning, no adequate distinction between curds and whey can be made to correctly get the curd sizes from its mask.

Differentiating themselves from the other techniques are the two AI methods, which score relatively high on both the precision as well as the recall (getting very close to the top right of the figure). Both outscor e the traditional segmentation methods, with the YOLOv8 model being the peak performer.

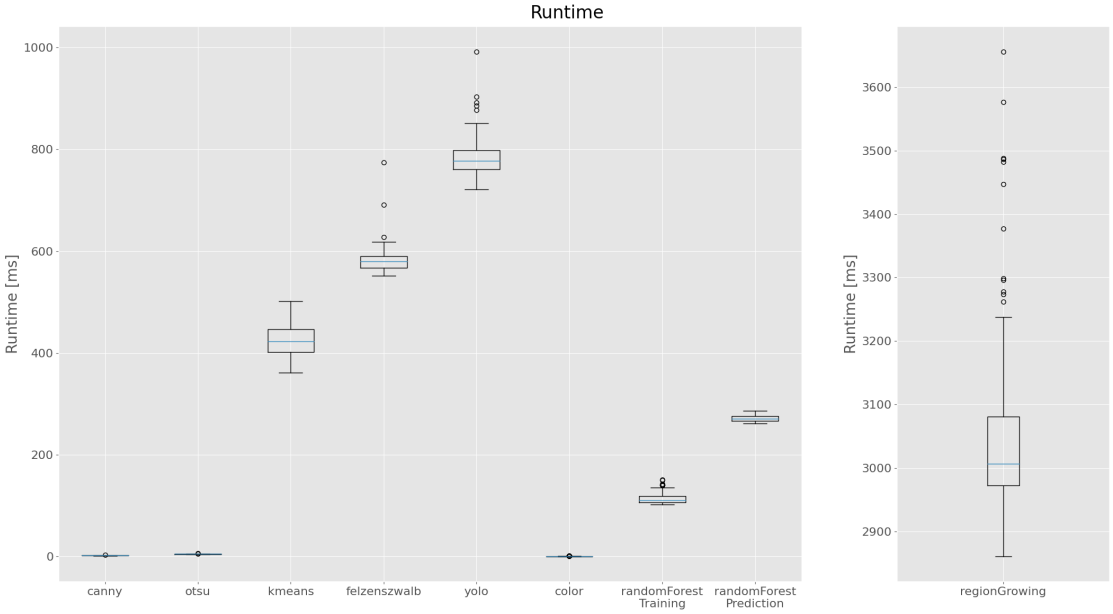


Figure 4.2: Runtime

Figure 4.2 shows the runtime of the methods. Each method was run 100 times to gather the runtime data. Pre-processing of the methods (i.e. possible image transformations from color to gray-scale) were not included in the runtime determination, to guarantee fair comparability. Ideally the segmentation methods will have a low runtime to be able to process as many frames per second and thus get as close as possible to real-time running. At the same time, ideally, a smaller spread of runtime is required to be able to process a consistent frame rate per second.

The first observation that can be made from these results is that the Canny, Otsu and color segmentation methods have an extremely low and consistent runtime. This is very much ideal to ensure as much information processing as possible during the cutting procedure of the cheese.

For the random forest method it stands out that the training time was lower than prediction. This can however be explained by the fact that the training occurs with only a random subset of image features, while at prediction the whole image is considered.

Additionally the figure shows that, while the YOLOv8 model has a high inference time on the CPU, it is still within reasonable runtime to run segmentation during the cutting process. The big outlier however in this data is the region growing method. It requires a runtime substantially higher than all other methods (and around three times as high as the YOLOv8 model), to such a degree that real-time running of this algorithm would significantly reduce the number of frames that can be investigated during the cutting process. Its very large spread, with fairly many outliers, is also not ideal to process a consistent number of frames per cutting process.

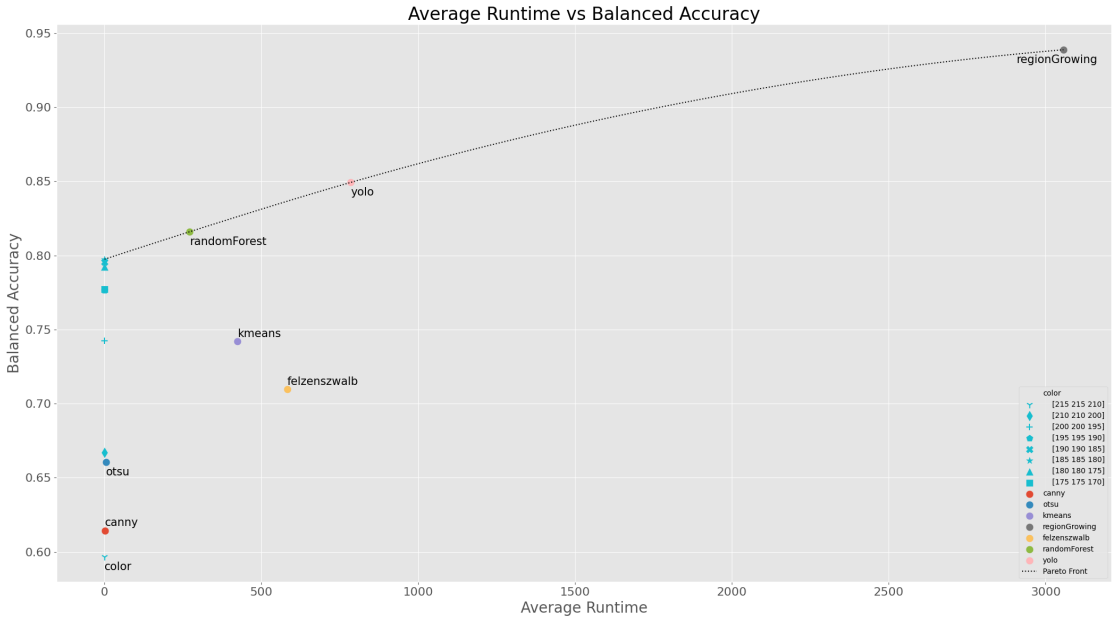


Figure 4.3: Runtime vs Balanced Accuracy

Figure 4.3 shows the balanced accuracy (as defined at the beginning of chapter 3) vs the runtime for every method. A noticeable first trend from the figure is that, in general, an increased runtime results in a higher balanced accuracy. The figure also includes a Pareto Front. A Pareto Front can be helpful to choose a solution that is Pareto Optimal (Lotov & Miettinen, 2008). As an example, considering the random forest segmentation in the figure, there is no other method that reaches a higher balanced accuracy while also having a lower runtime. Therefore the random forest method gets included in the Pareto Front.

From the figure it is also visible that while some methods (like Canny, Otsu or color segmentation) have the same very low runtime, many of them vary in accuracy. Based on this observation, and with the help of the Pareto Front, many methods drop out of contention, since there is an alternative method that performs better with a comparable, or lower, runtime. Felzenszwalb for example gets dominated by k-means, which in turn gets dominated by for example random forest.

Additionally it is clear that for color thresholding, while they all have the same runtime, a threshold value of $[185, 185, 180]$ scores the highest balanced accuracy, meaning that, even though further decreasing the threshold value resulted in an increasingly higher recall (as seen in figure 4.1), eventually overall performance decreases.

Additionally the region growing segmentation scores the highest balanced accuracy of all methods. It should however be emphasized that manual input (seed-point on a curd) was needed to achieve this result, combined with the fact that only one curd was used to assess its performance.

4.2. Requirements for cheese robot at Lely

While the above reviewed methods will all have a well-suited application, not all of them will be satisfactory for segmentation for the cheese robot at Lely. The main requirement for the cheeserobot is that with the segmentation, the mask is able to determine the curds with a high accuracy and has clear edges between the various curds. Additionally a low runtime is required to enable as much information processing as possible, as to get good insight into the cutting process. A method that is robust and can deal with changing conditions is required, so that human intervention is minimized and high uptime of the cheeserobot can be guaranteed.

From the metrics discussed in section 4.1, canny edge detection and otsu thresholding do not result in a balanced accuracy that is well enough to consistently segment curds from whey and, more importantly, accurately determine the size of the curds. The Felzenszwalb and kmeans methods perform better in terms of accuracy. However, from the Pareto Front shown in figure 4.3 it is visible that these methods get dominated in terms of accuracy and runtime.

By examining only the points on the Pareto Front, one can choose a solution that is Pareto Optimal and best suited for this purpose. The color thresholding with $[185, 185, 180]$ as threshold value is the quickest method with also adequate balanced accuracy. However, this method is very dependent on its tuning. This means that when (inevitably) the conditions near the cheese robot change (e.g. lighting conditions), the segmentation method should be tuned again, which hampers the desire for continuous cheese production without human intervention. Therefore a more robust method is needed that can withstand changing environmental conditions.

Segmentation with random forest can provide a solution, as this method can ”learn” to adapt in changing conditions by labeling data in such conditions. The same principle applies to the YOLOv8 model, which (being a deep learning model) can be robust to changing conditions, given that there is a robust labeled dataset (Shi et al., 2020). A key difference between both models, however, is that segmentation with random forest ”only” differentiates between classes of curd and whey. It would therefore require an additional algorithm to differentiate between the different instances, like the watershed algorithm (Beucher, 1979). The YOLOv8 model on the other hand does this at inference. With also a higher balanced accuracy, the YOLOv8 model seems very promising, as there is still room for improvement since the current model has only been trained on six images. While the inference time of this YOLOv8 model is on the higher side compared to the other methods, it is still within margins to cover multiple moments within the whole cutting process during cheesemaking.

The final point on the Pareto Front is the region growing segmentation. This method performs with the highest accuracy, although it also requires the highest runtime. While the runtime is high to such a degree that real-time running would no longer be possible, its biggest downside is that manual input is needed to perform segmentation of the curds. While this would be tolerable to do at a low frequency, the goal of the cheeserobot is to be a continuous process, making autonomous cheese. It is therefore not feasible to use this region growing method to perform segmentation on curds and whey in this context.

Based on these findings the YOLOv8 model is found to have a lot of potential, while being within the desired accuracy range and inference time. At the same time being able to differentiate between the instances without the need for an additional algorithm makes the YOLOv8 model very well suited for this application. Therefore the YOLOv8 model will be used further in this research.

Bibliography

Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*(6), 641–647. https://doi.org/10.1109/34.295913

Al-Amri, S. S., Kalyankar, N. V., et al. (2010). Image segmentation by using threshold techniques. *arXiv preprint arXiv:1005.4020*.

Aldalur, A., Bustamante, M. Á., & Barron, L. J. R. (2019). Characterization of curd grain size and shape by 2-dimensional image analysis during the cheesemaking process in artisanal sheep dairies. *Journal of Dairy Science*, *102*(2), 1083–1095. https://doi.org/https://doi.org/10.3168/jds.2018-15177

Bandyk, M. G., Gopireddy, D. R., Lall, C., Balaji, K., & Dolz, J. (2021). Mri and ct bladder segmentation from classical to deep learning based approaches: Current limitations and lessons. *Computers in Biology and Medicine*, *134*, 104472. https://doi.org/https://doi.org/10.1016/j.compbiomed.2021.104472

Basavaprasad, B., & Ravindra, S. H. (2014). A survey on traditional and graph theoretical techniques for image segmentation. *Int. J. Comput. Appl*, *975*, 8887.

Benallal, M., & Meunier, J. (2003). Real-time color segmentation of road signs. *CCECE 2003 - Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No.03CH37436)*, *3*, 1823–1826 vol.3. https://doi.org/10.1109/CCECE.2003.1226265

Beucher, S. (1979). Use of watersheds in contour detection. *Proc. Int. Workshop on Image Processing, Sept. 1979*, 17–21.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*, 197–227.

Bintsis, T. (2007). Quality of the brine. https://doi.org/10.1002/9780470995860.ch9

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

Camilus, K. S., & Govindan, V. (2012). A review on graph based segmentation. *International Journal of Image, Graphics and Signal Processing*, *4*(5), 1.

Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *PAMI-8*, 679–698. https://doi.org/10.1109/TPAMI.1986.4767851

Cashman, P., Kitney, R., Gariba, M., & Carter, M. (2002). Automated techniques for visualization and mapping of articular cartilage in mr images of the osteoarthritic knee: A base technique for the assessment of microdamage and submicro damage. *IEEE Transactions on NanoBioscience*, *1*(1), 42–51. https://doi.org/10.1109/TNB.2002.806916

Celenk, M., & de Haag, M. U. (1998). Optimal thresholding for color images. In E. R. Dougherty & J. T. Astola (Eds.), *Nonlinear image processing ix* (pp. 250–259). SPIE. https://doi.org/10.1117/12.304605

Chang, Y.-L., & Li, X. (1994). Adaptive image region-growing. *IEEE Transactions on Image Processing*, *3*(6), 868–872. https://doi.org/10.1109/83.336259

Chaple, G. N., Daruwala, R., & Gofane, M. S. (2015). Comparisions of robert, prewitt, sobel operator based edge detection methods for real time uses on fpga. *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, 1–4.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer vision – eccv 2018* (pp. 833–851). Springer International Publishing.

Chiu, K.-Y., & Lin, S.-F. (2005). Lane detection using color-based segmentation. *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, 706–711.

Cross, S., Henderson, J., & Dunkley, W. (1977). Losses and recovery of curd fines in cottage cheese manufacture. *Journal of Dairy Science*, *60*(11), 1820–1823. https://doi.org/https://doi.org/10.3168/jds.S0022-0302(77)84107-6

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. *Ensemble machine learning: Methods and applications*, 157–175.

Everard, C., O'Callaghan, D., Fagan, C., O'Donnell, C., Castillo, M., & Payne, F. (2007). Computer vision and color measurement techniques for inline monitoring of cheese curd syneresis. *Journal of Dairy Science*, *90*(7), 3162–3170. https://doi.org/https://doi.org/10.3168/jds.2006-872

Fagan, C. C., Castillo, M., Payne, F. A., O'Donnell, C. P., Leedy, M., & O'Callaghan, D. J. (2007). Novel online sensor technology for continuous monitoring of milk coagulation and whey separation in cheesemaking [PMID: 17854151]. *Journal of Agricultural and Food Chemistry*, *55*(22), 8836–8844. https://doi.org/10.1021/jf070807b

Farkye, N. Y. (2004). Cheese technology. *International Journal of Dairy Technology*, *57*(2-3), 91–98.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International journal of computer vision*, *59*, 167–181.

Fox, P., Guinee, T., Cogan, T., & McSweeney, P. (2017). *Fundamentals of cheese science*. https://doi.org/10.1007/978-1-4899-7681-9

Gan, H.-S., Ramlee, M. H., Wahab, A. A., Lee, Y.-S., & Shimizu, A. (2021). From classical to deep learning: Review on cartilage and bone segmentation techniques in knee osteoarthritis research. *Artificial Intelligence Review*, *54*(4), 2445–2494. https://doi.org/10.1007/s10462-020-09924-4

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: An overview. *arXiv preprint arXiv:2008.05756*.

Hartigan, J. A., Wong, M. A., et al. (1979). A k-means clustering algorithm. *Applied statistics*, *28*(1), 100–108.

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Hickey, C., Auty, M., Wilkinson, M., & Sheehan, J. (2015). The influence of cheese manufacture parameters on cheese microstructure, microbial localisation and their interactions during ripening: A review. *Trends in Food Science & Technology*, *41*(2), 135–148. https://doi.org/https://doi.org/10.1016/j.tifs.2014.10.006

Huth, P., DiRienzo, D., & Miller, G. (2006). Major scientific advances with dairy foods in nutrition and health. *Journal of Dairy Science*, *89*(4), 1207–1221. https://doi.org/https://doi.org/10.3168/jds.S0022-0302(06)72190-7

Iezzi, R., Locci, F., Ghiglietti, R., Belingheri, C., Francolino, S., & Mucchetti, G. (2012). Parmigiano reggiano and grana padano cheese curd grains size and distribution by image analysis. *LWT*, *47*(2), 380–385. https://doi.org/https://doi.org/10.1016/j.lwt.2012.01.035

Johnston, K. A., Dunlop, F. P., & Lawson, M. F. (1991). Effects of speed and duration of cutting in mechanized cheddar cheesemaking on curd particle size and yield. *Journal of Dairy Research*, *58*(3), 345–354. https://doi.org/10.1017/S0022029900029927

Lakshmanan, V., Görner, M., & Gillard, R. (2021). *Practical machine learning for computer vision*. " O'Reilly Media, Inc.".

Law, B. A., & Tamime, A. Y. (2010). *Technology of cheesemaking (2nd ed.)* Blackwell. https://doi.org/https://doi.org/10.1002/9781444323740

Li, C., Li, L., Geng, Y., Jiang, H., Cheng, M., Zhang, B., Ke, Z., Xu, X., & Chu, X. (2023). Yolov6 v3.0: A full-scale reloading. https://doi.org/10.48550/ARXIV.2301.05586

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755.

Lotov, A. V., & Miettinen, K. (2008). Visualizing the pareto frontier. *Multiobjective optimization*, *5252*, 213–243.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*(1167), 187–217.

Marth, E., & Steele, J. (2001). *Applied dairy microbiology*. CRC Press.

Mateo, M. J., O'Callaghan, D. J., Gowen, A. A., & O'Donnell, C. P. (2010). Evaluation of a vat wall-mounted image capture system using image processing techniques to monitor curd moisture during syneresis with temperature treatments. *Journal of Food Engineering*, *99*(3), 257–262. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2010.02.019

McBride, R. L., & Hall, C. (1979). Cheese grading versus consumer acceptability: An inevitable discrepancy [Copyright - Copyright Dairy Industry Association of Australia Jun 1979; Document feature - Tables; Graphs; ; Last updated - 2022-10-20; CODEN - AJDTAZ]. *Australian Journal*

of Dairy Technology, *34*(2), 66–68. https://www.proquest.com/scholarly-journals/cheese-grading-versus-consumer-acceptability/docview/304683629/se-2

McClure, S. B., Magill, C., Podrug, E., Moore, A. M. T., Harper, T. K., Culleton, B. J., Kennett, D. J., & Freeman, K. H. (2018). Fatty acid specific δ13c values reveal earliest mediterranean cheese production 7,200 years ago. *PLoS ONE*, *13*(9). https://doi.org/https://doi.org/10.1371/journal.pone.0202807

McSweeney, P. (2007). *Cheese problems solved*. Elsevier Science.

McSweeney, P., Fox, P., Cotter, P., & Everett, D. (2017). *Cheese: Chemistry, physics and microbiology: Fourth edition*.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968

Mo, Y., Wu, Y., Yang, X., Liu, F., & Liao, Y. (2022). Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, *493*, 626–646. https://doi.org/https://doi.org/10.1016/j.neucom.2022.01.005

Morris, O., Lee, M., & Constantinides, A. (1986). Graph theory for image analysis: An approach based on the shortest spanning tree. *Communications, Radar and Signal Processing, IEE Proceedings F*, *133*, 146–152. https://doi.org/10.1049/ip-f-1:19860025

Muir, D. D., & Hunter, E. A. (1992). Sensory evaluation of cheddar cheese: The relation of sensory properties to perception of maturity. *International Journal of Dairy Technology*, *45*(1), 23–30.

OpenCV. (2023a). *Canny edge detection*. Retrieved March 14, 2023, from https://docs.opencv.org/4.x/da/d22/tutorial_py_canny.html

OpenCV. (2023b). *Dilate*. Retrieved May 9, 2023, from https://docs.opencv.org/3.4/d4/d86/group__imgproc__filter.html#ga4ff0f3318642c4f469d0e11f242f3b6c

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, *9*(1), 62–66.

Padilla, R., Netto, S. L., & Da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. *2020 international conference on systems, signals and image processing (IWSSIP)*, 237–242.

Panthi, R. R., Kelly, A. L., McMahon, D. J., Dai, X., Vollmer, A. H., & Sheehan, J. J. (2019). Response surface methodology modeling of protein concentration, coagulum cut size, and set temperature on curd moisture loss kinetics during curd stirring. *Journal of Dairy Science*, *102*(6), 4989–5004. https://doi.org/https://doi.org/10.3168/jds.2018-15051

Papademas, P., & Bintsis, T. (2017). *Global cheesemaking technology: Cheese quality and characteristics*. https://doi.org/10.1002/9781119046165

Paperswithcode. (2023). *Semantic segmentation*. Retrieved March 13, 2023, from https://paperswithcode.com/task/semantic-segmentation

Piggott, J. R., & Mowat, R. G. (1991). Sensory aspects of maturation of cheddar cheese by descriptive analysis [Cited By :68]. *Journal of Sensory Studies*, *6*(1), 49–62. https://doi.org/10.1111/j.1745-459X.1991.tb00501.x

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. https://doi.org/10.48550/ARXIV.1506.01497

Renault, C., Gastaldi, E., Lagaude, A., Cuq, J., & De La Fuente, B. T. (1997). Mechanisms of syneresis in rennet curd without mechanical treatment. *Journal of Food Science*, *62*(5), 907–910. https://doi.org/https://doi.org/10.1111/j.1365-2621.1997.tb15004.x

Rogowska, J. (2009). Chapter 5 - overview and fundamentals of medical image segmentation. In I. N. Bankman (Ed.), *Handbook of medical image processing and analysis (second edition)* (Second Edition, pp. 73–90). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012373904-9.50013-1

ScienceLearningHub. (2012). *Creating different cheese characteristics*. Retrieved February 16, 2023, from https://www.sciencelearn.org.nz/resources/829-creating-different-cheese-characteristics

Shaaban, K. M., & Omar, N. M. (2009). Region-based deformable net for automatic color image segmentation [Special Section: Computer Vision Methods for Ambient Intelligence]. *Image and Vision Computing*, *27*(10), 1504–1514. https://doi.org/https://doi.org/10.1016/j.imavis.2009.02.003

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., & Shen, D. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for covid-19. *IEEE reviews in biomedical engineering*, *14*, 4–15.

Silanikove, N., Leitner, G., & Merin, U. (2015). The interrelationships between lactose intolerance and the modern dairy industry: Global perspectives in evolutional and historical backgrounds. *Nutrients*, *7*(9), 7312–31. https://doi.org/https://doi.org/10.3390/nu7095340

Soustre, Y., & Rayr, C. (2012). Special issue: Cheese. *Best Of*. https://www.fil-idf.org/wp-content/uploads/2016/06/SAY-CHEESE-Best-of-CNIEL-2012.pdf

Ultralytics. (2023a). *Ultralytics yolov8 docs*. Retrieved June 2, 2023, from https://docs.ultralytics.com/

Ultralytics. (2023b). *Ultralytics yolov8 docs - segment*. Retrieved June 2, 2023, from https://docs.ultralytics.com/tasks/segment/

Valous, N., & Sun, D.-W. (2012). 4 - image processing techniques for computer vision in the food and beverage industries. In D.-W. Sun (Ed.), *Computer vision technology in the food and beverage industries* (pp. 97–129). Woodhead Publishing. https://doi.org/https://doi.org/10.1533/9780857095770.1.97

Wang, W., Dai, J., Chen, Z., Huang, Z., Li, Z., Zhu, X., Hu, X., Lu, T., Lu, L., Li, H., et al. (2022). Internimage: Exploring large-scale vision foundation models with deformable convolutions. *arXiv preprint arXiv:2211.05778*.

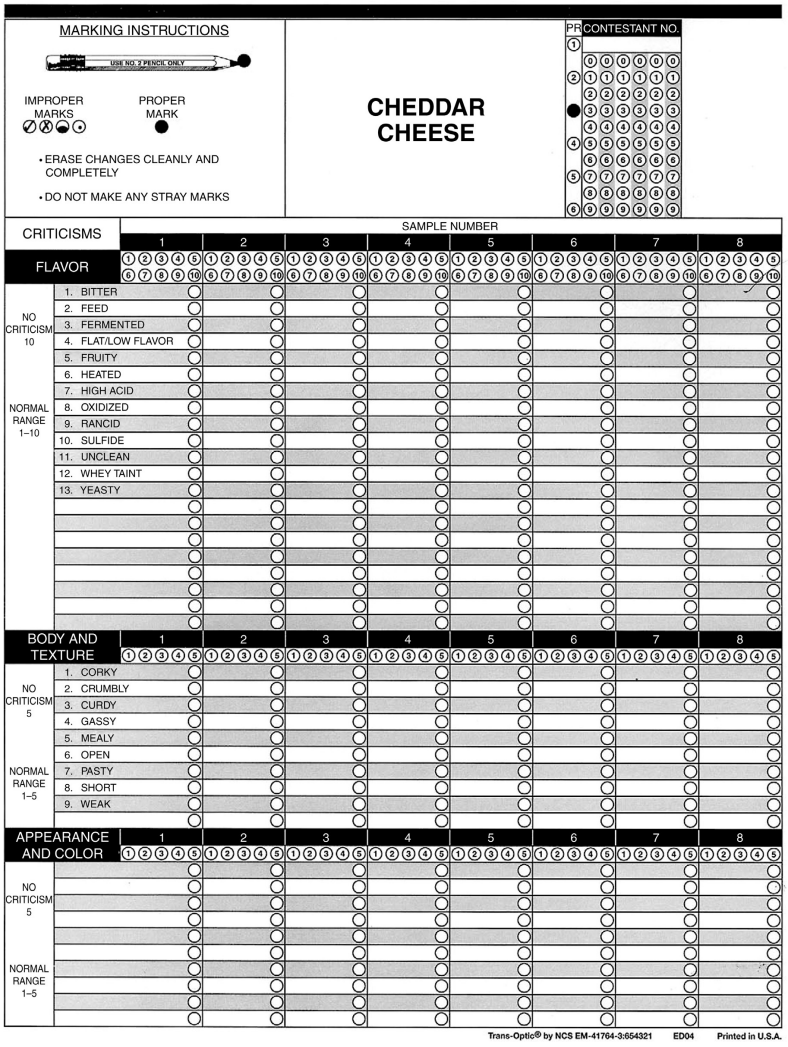
Wang, W., Bao, H., Dong, L., Bjorck, J., Peng, Z., Liu, Q., Aggarwal, K., Mohammed, O. K., Singhal, S., Som, S., & Wei, F. (2022). Image as a foreign language: Beit pretraining for all vision and vision-language tasks. https://doi.org/10.48550/ARXIV.2208.10442

Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. (2020). *Advances in deep learning*. Springer. https://doi.org/10.1007/978-981-13-6794-6

Whitehead, H. R., & Harkness, W. L. (1954). The influence of variations in cheese-making procedure on the expulsion of moisture from cheddar cheese curd [Copyright - Copyright Dairy Industry Association of Australia Jul-Sep 1954; Document feature - Tables; ; Last updated - 2014-05-18; CODEN - AJDTAZ]. *Australian Journal of Dairy Technology*, *9*(3), 103–107. https://www.proquest.com/scholarly-journals/influence-variations-cheese-making-procedure-on/docview/199492480/se-2

A

Defect Based Grading Form Cheese



MARKING INSTRUCTIONS

USE NO. 2 PENCIL ONLY

IMPROPER MARKS ⊘ ⊗ ◒ ⊙

PROPER MARK ●

- ERASE CHANGES CLEANLY AND COMPLETELY
- DO NOT MAKE ANY STRAY MARKS

CHEDDAR CHEESE

PR CONTESTANT NO.

CRITICISMS		SAMPLE NUMBER 1	2	3	4	5	6	7	8
FLAVOR									
NO CRITICISM 10	1. BITTER								
	2. FEED								
	3. FERMENTED								
	4. FLAT/LOW FLAVOR								
	5. FRUITY								
	6. HEATED								
	7. HIGH ACID								
NORMAL RANGE 1–10	8. OXIDIZED								
	9. RANCID								
	10. SULFIDE								
	11. UNCLEAN								
	12. WHEY TAINT								
	13. YEASTY								
BODY AND TEXTURE		1	2	3	4	5	6	7	8
NO CRITICISM 5	1. CORKY								
	2. CRUMBLY								
	3. CURDY								
	4. GASSY								
	5. MEALY								
NORMAL RANGE 1–5	6. OPEN								
	7. PASTY								
	8. SHORT								
	9. WEAK								
APPEARANCE AND COLOR		1	2	3	4	5	6	7	8
NO CRITICISM 5									
NORMAL RANGE 1–5									

Trans-Optic® by NCS EM-41764-3:654321 ED04 Printed in U.S.A.

Figure A.1: Defect Based Grading Form Cheese (McSweeney et al., 2017, p. 532)