A Large-Scale Dataset for Fish Segmentation and Classification

Oguzhan Ulucan, Diclehan Karakaya, Mehmet Turkan
Department of Electrical and Electronics Engineering
Izmir University of Economics, Izmir, Turkey
{oguzhan.ulucan, diclehan.karakaya, mehmet.turkan}@ieu.edu.tr

***Abstract*—Assessing the quality of seafood both in retail and during packaging at the production side must be carried out minutely in order to avoid spoilage which causes severe human health problems and also economic loss. Since the illnesses and decay in seafood presents distinct symptoms in different species, primarily the classification of species is required. In this field, the inadequacy of the current laborious and slow traditional methods can be overcome with systems based on machine learning and image processing, which present fast and precise results. In order design such systems, practical and suitable datasets are required. However, most of the publicly available datasets are not fit for the mentioned purpose. These datasets either contain images taken underwater or consist of seafood which is generally not (widely) consumed. In this study, a practical and large dataset containing nine distinct seafood widely consumed in the Aegean Region of Turkey is formed. Additionally, comprehensive experiments based on different classification approaches are performed to analyze the usability of this collected dataset. Experimental results demonstrate very promising outcomes; therefore, this dataset will be made publicly available for further investigations in this research domain.**

Keywords—Fish dataset, feature extraction, segmentation, classification, food quality assessment

I. Introduction

Seafood is the main ingredient of several dishes especially in seaside countries because of its taste and the nutrients it contains. Additionally, according to the Food and Agriculture Organization of the United Nations, the global fish consumption has risen above 20 kilograms a year [1]. Therefore, for the food industry, meeting the demand of consumers and supplying high quality products within a limited time has become more difficult. As a result, avoiding seafood spoilage in order to prevent economic loss and satisfy expectations of customers has gained more importance. Conventional methods of detecting spoilage and diseases, i.e., analyzing food samples in a laboratory, are laborious and time-consuming. Moreover, when experts try to detect spoilage through their past experience via sense of sight commonly, subjective decisions are taken which, for instance, may cause spoiled seafood to stay in the supermarket counter for sale. Developing an objective, fast and robust automated system can solve the problem of detecting spoilage and will lower the economic burden of vendors. In fact, fish breeds usually present different kinds of symptoms when they begin to develop a disease and start to spoil [2]. Hence, an automated classification of fish types is necessary before distinguishing fresh fish from spoiled ones and identifying diseases in these different species.

There are numerous studies present in the literature, which distinguish fish species through different algorithms using various datasets [2]–[4]. In feature-based fish classification systems, the feature extraction part is diverse and commonly used features are obtained from fish texture [2], [5], [6], size and shape [7]–[9]. For instance, the study in [9] for identifying ages of fish, otholith morphological features are extracted and their success rates are compared with sex, length and weight features through support vector machines (SVMs) based classifiers. Experimental results indicate that the classification accuracy reached up to 75% when all these features are combined. Besides feature-based methods, classification is also carried out using distinct techniques in fish identification tasks. In [10], a dataset containing six different fish types (grass carp, common carp, mori, rohu, silver carp, thala) is collected in Pakistan. Each class in the dataset contains three dominant features of fish types (body, scale and head) with different number of images. Convolutional neural networks (CNNs) are utilized to classify fish from their body images. Yet another study in [11], researchers developed a model for fish breeds recognition from underwater images via deep learning. While CNNs are selected to extract features from the Fish4Knowledge dataset [12], [13], SVMs and K-Nearest Neighbor (KNN) are employed for the classification task. The classification is performed successfully for both SVMs and KNN with accuracy rates of 98.32% and 98.79%, respectively.

An important challenge in fish classification studies is the lack of publicly available datasets containing commonly consumed fish image samples. Already present datasets contain fish images taken underwater and consist of fish species that are not usually consumed [10]–[15]. These datasets mostly provide data for marine biologists, scientists and researchers; hence they are not practically suitable for the food quality assessment problem. To the best of our knowledge, there is no publicly available dataset which contains seafood and fish sold in retail. This study fulfills the need of such a dataset containing image samples of eight different fish species (commonly consumed in the Aegean Region of Turkey) and shrimp which are collected from the fish counter of a supermarket. Also, comprehensive classification tasks are performed to analyze the usability of this dataset. In detail, the paper provides analysis results of the dataset through semantic segmentation and feature-based classification relying on gray-level co-occurrence matrices (GLCM), moments, bag-of-features (BoF) and CNNs features (CNNsF) via SVMs. The detailed analysis

978-1-7281-9136-2/20/$31.00 ©2020 IEEE

Figure 1: Example images from the collected dataset.

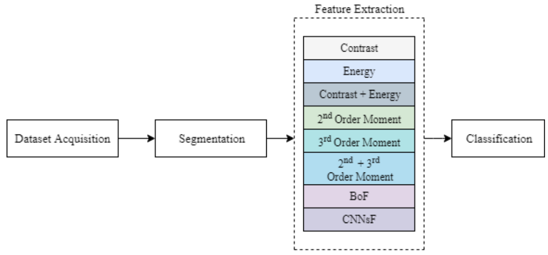

Figure 2: The block diagram of the experimental setup.

of this dataset will provide a guide light for future research, as it demonstrates success rates of the employed methods in seafood segmentation and classification tasks. Furthermore, the collected dataset and its ground-truth (manually extracted) segmentation masks will be publicly available for research purposes, which is one of the main contributions of this study. The rest of the paper is organized as follows. Section II presents the details of the collected dataset, and then demonstrates the experimental setup and results. Section III concludes this study with a brief summary.

II. Experimental Setup and Results

A. The Dataset

Images of nine different seafood types are collected from the fish counter of a supermarket. Two cameras are used in the dataset gathering process, a Kodak Easyshare Z650 and a Samsung ST60 with spatial resolutions of 2832×2128 and 1024×768 pixels, respectively. While 50 distinct fish images are collected per each of seven classes as follows: red mullet, gilt head bream, horse mackerel, sea bass, red sea bream, black sea sprat and striped red mullet, 30 distinct images are captured for trout and shrimp. All fish in the image acquisition process are fresh, and they are positioned in various displacements and angles but lighting conditions do not change significantly. Lastly, instead of a clean white background, a blue and noisy background is preferred in order to make the dataset usable in studies with real-life problems. Example images from the collected dataset are illustrated in Fig. 1. Furthermore, the sample images of all nine classes are resized to 590×445 pixels by nearly preserving their aspect ratio. Then, these samples are passed through an augmentation algorithm where each image is rotated with non-repeated random angles and they are reflected. For each seafood type, 1000 images are finally obtained for the construction of the dataset.

Several experiments are carried out to analyze the performance of the collected dataset. The basic procedure is based on fish segmentation and then feature-based fish classification (relying on different feature types) using SVMs based classifiers. The block diagram of the experimental setup is given in Fig. 2 and the obtained experimental results are reported in the remaining part of this section.

B. Fish Segmentation

In the first set of experiments, the ground-truth segmentation masks of all seafood images are extracted manually by a human operator. Then, morphological operators are used to smooth these masks and to obtain finer shaped binary labels. During this process, an erosion operation with the shape of diamond and distance of 8 is adopted. Later, a sphere shaped dilation operator with the distance of 25 is applied to all masks to obtain final ground-truth masks.

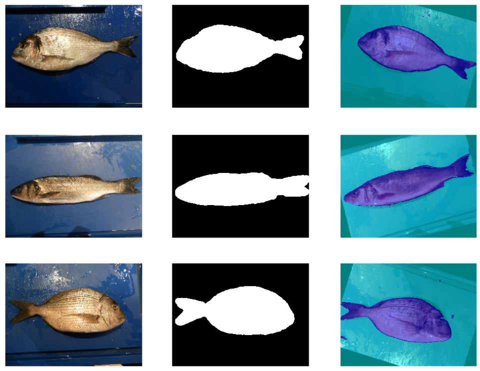

Figure 3: (Left-to-right) Fish image samples; the ground-truth segmentation masks; segmentation results of SegNet.

After obtaining ground-truth masks manually, the fish to be separated automatically from its background through a semantic segmentation algorithm called SegNet [16], which is a neural network containing ten layers. The whole data is divided as of 70% training set and of 30% test set. For the two classes (in this case, fish and background), the filter size and the number of filters were determined as 3×3 and 64, respectively. The maximum epoch number is taken as 10 and the mini-batch size is chosen as 8. It is observed that the background contains a significant amount of pixels more than the fish does. Therefore, the weights of the classes are calculated by using inverse frequency weighting and the last layer of the network is updated. At the end, SegNet reached an average of 98.01% and 88.69% training and test accuracy rates, respectively. Figure 3 demonstrates some examples of the ground-truth masks and SegNet results on the augmented dataset. In addition, Table I reports the detailed segmentation rates obtained through Jaccard similarity index (percentages).

C. Feature Extraction

After automatic segmentation, all seafood types were classified via the extracted features from segmented images. The information provided to the classifier are obtained through four different feature extraction methods as follows.

	Train Accuracy	Test Accuracy
Gilt Head Bream	97.12	96.82
Red Sea Bream	98.21	91.84
Sea Bass	95.69	82.98
Red Mullet	99.02	90.32
Horse Mackerel	98.90	86.97
Black Sea Sprat	97.52	89.66
Striped Red Mullet	99.24	89.60
Trout	97.32	80.45
Shrimp	99.06	89.59
Average	98.01	88.69

Table I: SegNet segmentation results in percentages.

1) GLCM: GLCM presents how frequent distinct gray-scale intensities of pixels appear in an image [17]. In this study, "contrast" and "energy" features are extracted through GLCM. While the contrast feature exploits the contrast between each pixel and its neighbor, the energy feature computes the sum of squared elements. The contrast feature can capture color variations of different fish species, hence the effect of distinct colors is aimed to be observed in the classification process. The energy feature on the other hand is chosen because of its proven efficiency in food quality assessment tasks [18]. These two features are fed to SVMs not only separately but also as a one concatenated feature.

2) Moments: The second order moment (variance) and the third order moment (skewness) are extracted to observe the distributions in the data and their effects in the classification success rates. Similar to GLCM features, the concatenation of the second and third order moments are performed to observe whether their combination results in improved test accuracy rates in the classification process.

3) BoF: The BoF algorithm is an adoption of the Bag of Words for image processing [19]. This feature extraction technique uses the SURF method to extract discrete pixel-based features to create a visual vocabulary [20]. After extracting all features, the weak ones are eliminated by using K-means clustering to obtain the final version of the visual vocabulary. Since it takes a great number of features into account, BoF is employed in this study.

4) CNNsF: CNNs are inspired by the human visual system and the biological structure of the brain and specifically designed for image processing [18]. CNNs mainly consist of four main layers; an input layer, a convolutional layer, a pooling layer and a fully connected layer [21]. The number and order of layers are selected according to the difficulty of the problem at hand. More complex problems usually require a higher number of layers [22]. In this study, CNNsF is adopted to extract various features such as edges, blobs and small details [23]. Since training an end-to-end network requires high computational time, AlexNet is employed as a pre-trained network. The epoch number is selected as 10 and the mini-batch size is determined as 8.

D. Classification Results

Since firstly introduced in 1979 as a supervised binary classifier, SVMs [24] have been widely used in classification tasks. They are preferred as classifiers in this study because of their success in food assessment studies, e.g., [18], [22], [25]. In this set of experiments, one-versus-all SVMs classifiers are designed for analyzing the use of the extracted features. While randomly 70% of the images are used for training, 30% are employed to test the algorithm. 10-fold cross-validation is used along with the SVMs and all feature vectors are normalized and formed with the same dimensions. Additionally, all experiments have been repeated five times with random initialization and average values are provided in the experimental results.

For the experiments, eight independent SVMs are trained. The training and test results are reported in Table II and Table III. The lowest average training and test accuracy rates are obtained by using BoF (84.36% and 81.55%, respectively). Actually, the BoF algorithm chooses the strongest features among others and it sometimes fails to select the suitable features for the given task when the images contain very similar texture and color. On the other side, the best training and test results are observed when the GLCM contrast feature is adopted in the SVMs classifier. A mean accuracy of 98.74% is achieved in the training process, which is almost 1% higher than the result obtained through CNNs-based features. For the average test accuracy, the difference between the contrast feature and CNNsF increased to more than 4%. Considering that CNNs are specifically designed for successful vision-based applications, it can be indicated here that the adoption of the GLCM contrast feature is a good choice for this dataset.

The concatenation of different features may result in different effects on the accuracy for distinct data as observed in this study. For instance, while the combination of contrast and energy features increases both training and test accuracy rates for red sea bream, it decreases the success of the system for black sea sprat. While using the energy feature alone, it produces a test accuracy of 97.15% for trout; however a success rate of 97.28% is achieved when it is combined with the contrast feature. Another point worth to note that is, employing the contrast and energy features individually leads to a higher accuracy on average than using their concatenation.

According to the classification results obtained through statistical features, it is observed that the second order moment leads to better accuracy rates on average. For horse mackerel and trout, it produces higher accuracy rates than the GLCM contrast feature in training. Another observation is that the concatenation of second and third order moments results in a lower average accuracy than using these features alone.

Lastly, the adoption of CNNsF in the SVMs algorithm leads to significantly higher accuracy rates than BoF and it closely follows the success rates of the GLCM contrast and energy features in the training process. Also for shrimp, low and high level features obtained through CNNsF produce the best training and test accuracy rates.

III. Conclusion

In the last two decades, the mutual employment of computer vision and machine learning has a strong impact and benefit in the food industry and its potential is recognized by many industrial companies up to date. Automated systems for food quality assessment have significantly decreased the time-consumption for this task while increasing the accuracy of spoilage detection. Similarly for the seafood, distinguishing the fish types is important for detecting the decay and illnesses. Unfortunately, there is a lack of publicly available datasets for

	Contrast	Energy	Contrast+Energy	2^{nd} Order Moment	3^{rd} Order Moment	2^{nd}+ 3^{rd} Order Moment	BoF	CNNsF
Gilt Head Bream	99.56	98.44	98.09	88.93	91.85	90.52	86.06	96.31
Red Sea Bream	98.85	99.33	99.39	98.22	89.00	89.89	66.61	97.86
Sea Bass	97.30	98.07	96.74	95.59	95.22	90.09	78.75	97.31
Red Mullet	99.56	99.19	98.81	89.78	96.67	91.52	92.14	97.62
Horse Mackerel	98.81	99.22	97.93	98.96	95.96	92.09	83.04	98.50
Black Sea Sprat	99.00	97.81	96.85	97.89	95.37	91.48	95.18	98.45
Striped Red Mullet	98.48	98.89	98.17	91.67	88.93	88.89	67.11	95.71
Trout	99.48	97.85	97.98	99.74	94.04	93.50	97.39	98.75
Shrimp	97.63	98.22	97.37	93.59	93.70	92.02	92.93	98.74
Average	98.74	98.56	97.96	94.93	93.42	91.11	84.36	97.81

Table II: Feature-based classification results for training of SVMs in percentages.

	Contrast	Energy	Contrast+Energy	2^{nd} Order Moment	3^{rd} Order Moment	2^{nd}+ 3^{rd} Order Moment	BoF	CNNsF
Gilt Head Bream	98.00	97.26	96.41	87.93	90.11	89.04	81.67	93.07
Red Sea Bream	97.41	98.67	98.81	95.04	86.15	88.48	64.58	96.17
Sea Bass	95.56	96.37	95.24	91.93	91.30	89.19	75.00	92.08
Red Mullet	98.89	98.44	98.06	86.85	92.93	90.46	87.25	89.90
Horse Mackerel	97.41	98.15	96.74	96.89	92.85	90.81	80.33	92.74
Black Sea Sprat	97.85	96.41	95.83	89.37	92.37	90.70	93.33	96.07
Striped Red Mullet	97.33	97.85	97.11	87.67	86.93	86.72	65.00	86.66
Trout	99.19	97.15	97.28	98.15	91.78	91.07	95.83	95.00
Shrimp	97.15	96.96	95.57	89.00	90.52	89.67	90.95	97.56
Average	97.64	97.47	96.78	91.43	90.55	89.57	81.55	93.25

Table III: Feature-based classification results for testing of SVMs in percentages.

analyzing the seafood quality in supermarket retail sections. In this study, a dataset from nine widely consumed seafood in the Aegean Region of Turkey is constructed while considering practical real-life constraints such as noisy backgrounds and distinct exposures. During the analysis of the dataset, ground-truth labels are obtained manually, semantic segmentation and feature-based classification tasks have been tested. Promising results are obtained for the employment of the collected dataset, hence it will be publicly available for further research.

REFERENCES

[1] FAO, "Global per capita fish consumption rises above 20 kilograms a year," 2016.

[2] J. Hu, D. Li, Q. Duan, Y. Han, G. Chen, and X. Si, "Fish species classification by color, texture and multi-class support vector machine using computer vision," *Comp. Electr. Agricult.*, vol. 88, pp. 133–140, 2012.

[3] R. Nijhawan, "A hybrid deep learning approach for automatic fish classification," *Proc. Emerging Trends Inf. Tech.*, p. 427, 2019.

[4] S. Hasija, M. J. Buragohain, and S. Indu, "Fish species classification using graph embedding discriminant analysis," in *Int. Conf. Mach. Vis. Inf. Tech.*, 2017.

[5] C. Spampinato, D. Giordano, R. D. Salvo, Y.-H. J. Chen-Burger, R. B. Fisher, and G. Nadarajan, "Automatic fish classification for underwater species behavior understanding," in *Proc. ACM Int. W. Anal. Retriev. Tracked Events Motion Imagery Streams*, 2010.

[6] C. Pornpanomchai, B. Lurstwut, P. Leerasakultham, and W. Kitiyanan, "Shape and texture based fish image recognition system," *Kasetsart J.-Natural Sci.*, vol. 47(4), pp. 624–634, 2013.

[7] S. O. Ogunlana, O. Olabode, S. A. A. Oluwadare, and G. B. Iwasokun, "Fish classification using support vector machine," *African J. Comput. & ICT*, vol. 8(2), pp. 75–82, 2015.

[8] R. Larsen, H. Olafsdottir, and B. K. Ersboll, "Shape and texture based classification of fish species," in *Scandinavian Conf. Image Anal.*, 2009.

[9] S. Bermejo, "Fish age classification based on length, weight, sex and otolith morphological features," *Fisheries Res.*, vol. 84(2), pp. 270–274, 2007.

[10] S. Z. H. Shah, H. T. Rauf, M. IkramUllah, M. S. Khalid, M. Farooq, M. Fatima, and S. A. C. Bukhari, "Fish-Pak: fish species dataset from Pakistan for visual features based classification," *Data Brief*, vol. 27, p. 104565, 2019.

[11] B. V. Deep and R. Dash, "Underwater fish species recognition using deep learning techniques," in *IEEE Int. Conf. Signal Process. Integrated Netw.*, 2019.

[12] B. J. Boom, P. X. Huang, J. He, and R. B. Fisher, "Supporting ground-truth annotation of image datasets using clustering," in *Proc. Int. Conf. Patt. Recog.*, 2012.

[13] B. J. Boom, J. He, S. Palazzo, P. X. Huang, C. Beyan, H.-M. Chou, F.-P. Lin, C. Spampinato, and R. B. Fisher, "A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage," *Ecol. Inform.*, vol. 23, pp. 83–97, 2014.

[14] G. Cutter, K. Stierhoff, and J. Zeng, "Automated detection of rockfish in unconstrained underwater videos using haar cascades and a new image dataset: labeled fishes in the wild," in *IEEE Winter Appli. Comp. Vis. W.*, 2015.

[15] B. J. Boom *et al.*, "Long-term underwater camera surveillance for monitoring and analysis of fish populations," *Proc. Vis. Observation Anal. Animal Insect Behav.*, 2012.

[16] (Accessed May 6, 2020) SegNet. https://uk.mathworks.com/help/vision/ug/semantic-segmentation-examples.html.

[17] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cyber.*, vol. SMC-3(6), pp. 610–621, 1973.

[18] O. Ulucan, D. Karakaya, and M. Turkan, "Meat quality assessment based on deep learning," in *Innov. Intell. Syst. Appli. Conf.*, 2019.

[19] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *W. Stat. Learn. Comp. Vis.*, 2004.

[20] E. Oyallon and J. Rabin, "An analysis of the SURF method," *Image Process. On Line*, vol. 5, pp. 176–218, 2015.

[21] D. Karakaya, O. Ulucan, and M. Turkan, "Electronic nose and its applications: a survey," *Int. J. Automat. Comput.*, vol. 17, pp. 179–209, 2020.

[22] D. Karakaya, O. Ulucan, and M. Turkan, "A comparative analysis on fruit freshness classification," in *Innov. Intell. Syst. Appli. Conf.*, 2019.

[23] (Accessed May 6, 2020) Model. https://ww2.mathworks.cn/help/vision/examples/image-category-classification-using-deep-learning.html.

[24] D. Boswell, "Introduction to support vector machines," *Dept. Comp. Sci. Eng., Univ. Calif. San Diego*, 2002.

[25] K. S. Tolentino, J. W. F. Orillo, P. D. Aguacito, E. J. M. Colango, J. R. H. Malit, J. T. G. Marcelino, A. C. Nadora, and A. J. D. Odeza, "Fish freshness determination through support vector machine," *J. Telecommun. Electron. Comput. Eng.*, vol. 9, pp. 139–143, 2017.