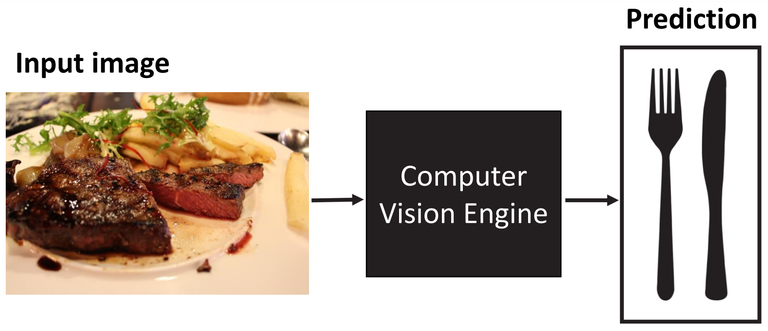
Understanding Food Images to Recommend Utensils During Meals

Francesco Ragusa, Antonino Furnari, Giovanni Maria Farinella
University of Catania – Department of Mathematics and Computer Science
francescoragusa@outlook.com, {furnari, gfarinella}@dmi.unict.it

IMAGE PROCESSING LABORATORY

UNIVERSITÀ degli STUDI di CATANIA

The Investigated Problem



Can we train a computer vision system to recognize the eating utensils to be used during a meal in order to help patients with dementia diseases in reminding how to eat food? Given a food image, the computer vision engine should be able to predict which utensils are to be used to consume the meal.

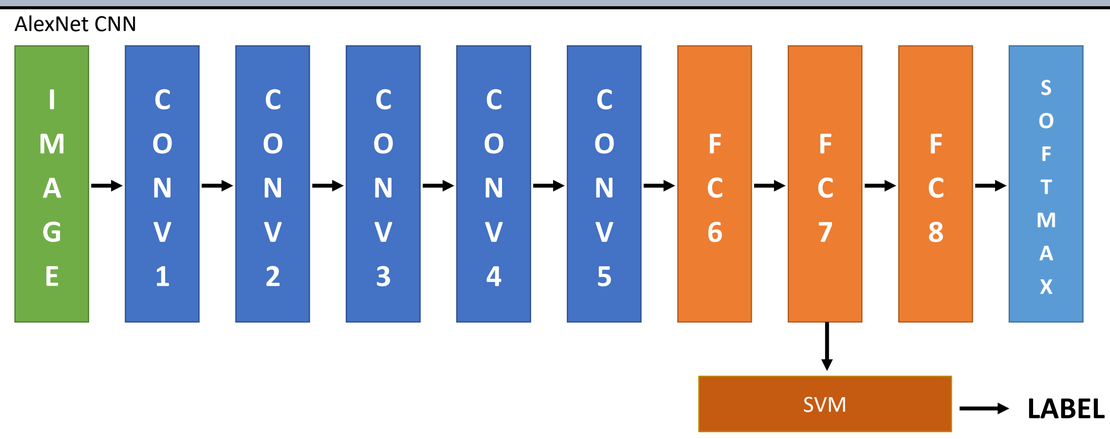
References

[1] Ragusa F., Tomaselli V., Furnari A., Battiato S., Farinella G. M., Food vs Non-Food Classification, 2nd International Workshop on Multimedia Assisted Dietary Management (MADiMa) in Conjunction with ACM Multimedia, Amsterdam, 2016.

[2] Furnari A., Farinella G.M., Battiato S Recognizing Personal Locations From Egocentric Videos, IEEE Transactions on Human-Machine Systems, Vol.47, pp. 6-18, 2017.

[3] Farinella G. M., Allegra D., Moltisanti M., Stanco F., Battiato S., Retrieval and classification of food images, Computers in Biology and Medicine, 77, 2339, 2016.

The Considered Approach



We choose to extract features form the FC7 layer since such activations are believed to have a high semantic relevance but are more general than the 1000 features of the FC8 layer which are to be considered as class-related scores. Classification is performed using a multiclass SVM classifier.

Dataset

Images belonging to the UNICT-FD1200 dataset.
Each row corresponds to a specific class: 1) *Chopsticks*, 2) *Fork*, 3) *Fork and Knife*, 4) *Hands*, 5) *Spoon*

The UNICT-FD1200 dataset is available at http://iplab.dmi.unict.it/UNICT-FD1200

Experiments and Results

We consider the UNICT-FD1200 dataset. To perform evaluation, the dataset has been randomly divided into three balanced non-overlapping subsets. This simple pipeline composed by CNN features and SVM classifier has obtained an accuracy of **86,27%**. The three different splits allow to obtain three independent training set/test set pairs.
Once FC7 features are extracted for all images in the dataset, the SVM classifier is trained and tested considering the three different split.

	Chopsticks	Fork	Fork and Knife	Hands	Spoon
Chopsticks	**60,87%**	6,52%	28,26%	2,17%	2,17%
Fork	0,00%	**86,47%**	11,88%	0,82%	0,82%
Fork and Knife	0,27%	17,00%	**92,88%**	2,19%	0,00%
Hands	0,84%	2,52%	14,28%	**82,35%**	0,00%
Spoon	1,78%	7,14%	17,85%	1,78%	**72,43%**