Evaluation of Convolutional Neural Networks for Raw Food Texture Classification under Variations of Lighting Conditions

Carolina Toledo Ferraz*, Tamiris T. N. Borges†, Adriane Cavichiolli‡, Adilson Gonzaga‡ and José Hiroki Saito*
*UNIFACCAMP, Campo Limpo Paulista/SP
caroltoledoferraz, saitojosehiroki@gmail.com
† Federal Institute of São Paulo (IFSP), Araraquara/SP
tamirisnegri@ifsp.edu.br
‡University of São Paulo
adriane.cavichiolli@usp.br, agonzaga@sc.usp.br

***Abstract*—This work is a preliminary evaluation of convolutional neural networks (CNN) applied to food texture classification, particularly when the texture is subject to changes in the lighting conditions. Four previously published CNN architectures (Alexnet, Resnet 18, Resnet 34 and Resnet 50) are investigated and compared to local descriptors designed specifically for this task. Although preliminary results indicate that the investigated CNN are outperformed by the descriptors, further analysis are required to investigate the impact of the experimental design adopted in this work-in-progress; especially in regard to the number of training samples and CNN configuration.**

I. Introduction

The food industry has quality standards that require constant assessment to avoid endangering consumer's health. Quality control is commonly performed manually by experienced workers; however, such process is laborious, costly and subjective. Thus, automating this task is of great interest to this field.

Martinel et al. [1] evaluate different texture filter banks for automatic food image recognition, motivated by a lack of studies showing which texture features are more suitable for this task. Ragusa et al. [2] explore food vs non-food classification using different deep-learning-based representations and classification methods. Pouladzadeh and Shirmohammadi [3] proposed a deep-learning method for detecting multiple food items from pictures taken using mobile devices. Zareiforoush et al. [4] presented a study of qualitative grading of milled rice grains using a machine vision system combined with metaheuristic classification approaches.

One of the challenges faced by the cited works is the variation in the visual information captured by the devices, due to changes in the view-angle, illumination color and direction. Variations in the light intensity, direction, and temperature may change the color of the observed texture and interfere in the capacity of the computer vision methods. Recent works proposed new color texture descriptors especially designed for image description under variations of lighting conditions [5], [6]. According to the results presented in [6], [7], changes in the lighting direction are the most challenging. Other works explored the use of convolutional neural networks (CNN) for colored texture classification [8] with adaptive changes in the networks for better image recognition.

This work proposes the investigation of four known CNN architectures - Alexnet, Resnet18, Resnet34 and Resnet50 - for food classification under the variation of light intensity, direction and temperature. The CNN performances were compared to four color texture descriptors - Extended Color Local Mapped Pattern (ECLMP) [7], Color Intensity Local Mapped Pattern (CILMP) [6], Opponent Color Local Mapped Pattern (OCLMP) [9] and Opponent Color Local Binary Pattern (OCLBP) [10].

II. Convolutional Neural Networks and Texture Descriptors for texture analysis

A. Convolutional Neural Networks (CNN)

Currently, CNN are the most popular deep-learning network models. Among the CNN architectures, AlexNet [11] and Residual Network (ResNet) [12] stand out for the excellent performance in image classification tasks. We choose AlexNet, ResNet18, ResNet34 and ResNet50 to evaluate the CNN performance for food classification under varying illumination. AlexNet is composed of 8 layers, 5 of them convolutional, and 3 totally connected. The ResNet architecture is deeper than AlexNet but with not-so-complex convolutional filters. The number in the nomenclature represents the number of layers.

B. Descriptors

To compare the performance of the CNN, we selected four descriptors designed to describe color images under varying illumination: ECLMP [7], CILMP [6], OCLMP [9] and OCLBP [10]. The parameter tunning for ECLMP, CILMP and OCLMP, and the experimental setup for the four descriptors were performed as described in [7].

III. Material and Method

In this work, we analyze the performance of the CNN described in Section II-A in the texture classification scenario and compare them against color texture descriptors (Section II-B). For this task, we considered sets of images from the Raw Food Texture Database (RawFooT) [13], which contains 68 textures of raw food such as meat, fish, cereals, fruit, etc. Each texture of the database was captured under 46 different lighting conditions to evaluate the robustness of computational methods to variations in the illumination.

We analyzed three lighting conditions: intensity, direction and temperature, as detailed below and presented in Fig. 1.

1) Light intensity: images acquired under **4** intensity levels of simulated daylight at 6500K (D65, I=100%, 75%, 50%, and 25%).
2) Light direction: images taken under simulated daylight at 6500K and **9** different light directions (D65, $\theta = 24$, 30, 36, 42, 48, 54, 60, 66, and 90 degrees).
3) Daylight temperature: images taken under simulated daylight at **12** temperatures varying from 4000K to 9500K (D40, D45, D50, . . . , D95).

The original 800×800 pixels textures were divided into 16 non-overlapping samples of 200×200 pixels. The total number of samples used in the experiments was 27200.

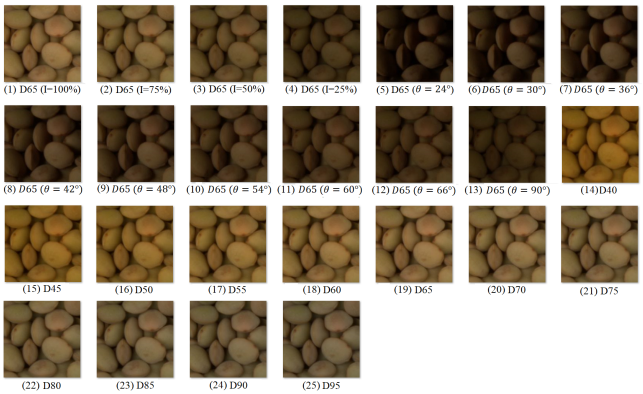

Fig. 1. Example of one of the textures (lentils) imaged under the 25 lighting conditions.

To perform the CNN training, the following parameters were used: a) training epochs: 1000; b) learning rate: 0.01; c) momentum: 0.9 and d) BatchSize: 32. The experiments were conducted on a CPU featuring an i7 processor, with 16GB RAM and a GeForce GTX Titan XP.

IV. Experiments

We divided the experiments into three groups, according to the lighting variation type: A) Light intensity, B) Light direction and C) Daylight temperature. For each group, two experiments were conducted, as described in the following sections.

A. Light Intensity

In the first experiment (A1) the CNN training was performed using images acquired under the maximum lighting intensity (100%). In this process, 12 samples per class were used for training and 4 samples per class were used for validation. The trained model was tested with images acquired under 75% (16 samples per class), 50% (16 samples per class) and 25% (16 samples per class) of the maximum illumination intensity. For the experiments performed by the texture descriptors, the samples acquired under maximum intensity were also used as training set, and the samples taken under 75%, 50% and 25% of the maximum intensity were used as test sets. Results are shown in Table I.

For the second experiment (A2), we selected 8 samples per class from each of the four intensities (32 samples) to train the CNN (23 samples for training and 9 for validation). The other 8 samples per class of each intensity were used to test the model. The exact same samples used for the CNN training and validation were used as the training set for the descriptor. Results are shown in Table II.

TABLE I
A1 - Classification accuracy (%) obtained by the models trained with images acquired at maximum intensity (100%)

Models	Intensity of the test samples			Avg. Accuracy
	25%	50%	75%	
Alexnet	0.41	19.94	61.67	27.34
Resnet18	0.51	24.90	53.40	26.27
Resnet34	0.59	22.24	62.40	**28.41**
Resnet50	0.18	18.01	65.80	27.99
ECLMP	77.57	94.85	97.24	**89.89**
CILMP	60.11	90.99	97.61	82.90
OCLMP	73.16	92.10	96.14	87.13
OCLBP	80.70	91.18	94.12	88.66

TABLE II
A2 - Classification accuracy (%) obtained by the models trained with four intensity levels

Models	Intensity of the test samples				Avg. Accuracy
	25%	50%	75%	100%	
Alexnet	88.78	93.19	92.09	91.36	91.35
Resnet18	86.02	92.46	92.83	94.66	**91.49**
Resnet34	85.29	91.91	91.17	92.83	90.30
Resnet50	84.55	91.36	90.25	89.33	88.87
ECLMP	98.53	99.26	99.45	99.45	**99.17**
CILMP	97.06	97.43	98.16	97.98	97.66
OCLMP	98.16	98.16	98.53	98.53	98.35
OCLBP	97.24	96.51	95.59	95.77	96.28

Table I shows that the highest average accuracy was achieved by the ECLMP (89.89%) when only one illumination intensity was considered during the training process. As expected, the classification accuracy drops as the difference between the intensity levels of training and test samples increases. Table II shows that the ECLMP also achieved the highest accuracy (99.17%) when all the illumination intensities are considered for training. Tables I and II show that the CNN are outperformed by all the descriptors in both experiments.

B. Light direction

In the first experiment (B1) we used images acquired under the lighting direction of $\theta = 24^\circ$ to train the CNN. In this process, 12 images per class were used for training and 4 images per class were used for validation. The trained model was tested with the following sets of images: $\theta = 30, 36, 42, 48, 54, 60, 66$ and 90 degrees. Each set contains 16 samples per class. For the experiments performed by the texture descriptors, the samples acquired under lighting direction of 24 degrees were used as training set, and the samples taken under the other directions were used as test samples. Results are shown in Table III.

For the second experiment (B2), we selected 8 samples per class from each of the nine light directions (72 samples per class) to train the CNN architectures (52 for training and 20 for validation). The other 8 samples per class of each direction were used to test the model. In the texture descriptor experiments, the same samples as used to train the CNN were used as training set, and the test sets were the same as used to test the CNN. Results are shown in Table IV.

Table III shows that the accuracy of the best descriptor in the B1 experiment (OCLBP, 68.17%) is 30% higher than the best CNN. Table IV shows that the highest accuracy in the B2 experiment is obtained by the ECLMP descriptor (98.72%). Again, the CNN were outperformed by the descriptors.

C. Daylight temperature

In the first experiment (C1) we used the images acquired under simulated daylight at 4000K (D40) to train the CNN. In this process, 12 images per class were used for training and 4 images per class were used for validation. After trained, the model was tested on the images taken at the eleven remaining temperatures ranging from 4500K to 9500K, considering 16 samples per class for each temperature. For the experiments performed by the texture descriptors, the samples acquired under simulated daylight at 4000K were also used as training set, and the samples taken under the other temperatures were used as test sets. Results are shown in Table V.

For the second experiment (C2), we selected 8 samples per class from each of the twelve daylight temperature (96 samples per class) to train the CNN architectures (69 for training and 27 for validation). The other 8 samples per class of each temperature were used to test the model. For the experiments performed by the texture descriptors, the same samples used to train the CNN were used as training set; the test sets were the same as used to test the CNN. Results are shown in Table VI.

Table V shows that the highest average accuracy in the experiment C1 was achieved by the CILMP (97.00%), while in the C2 experiment the best accuracy was achieved by the ECLMP (99.19%). Once more, the CNN were outperformed by the texture descriptors in the classification task.

V. Discussions and Conclusions

This work-in-progress presented the preliminary work on the application of CNN to the classification of food texture, and its comparison against texture descriptors.

Tables I, III and V make evident that the CNN performed considerably worse than the investigated descriptors when the model was trained at a lighting condition different then the present in the testing set.

Tables II, IV and VI show that even though the performance of the CNN increased drastically by including samples from all illumination conditions to the training set, the CNN are still outperformed by the descriptors in all experiments.

It is important to highlight that, even though the preliminary results indicate that CNN perform worse than local descriptors in the task of food classification, further analysis is necessary to support any claims. Processing time is an important factor for this work in development, because generating the feature descriptors takes time, which is impracticable for applications that need to be executed in real time.

This work-in-progress will proceed with a complete analysis about the impact of using a low number of training samples to the convergence of the CNN, it will also include further investigation about CNN configurations appropriate for this application.

Acknowledgment

The authors would like to thank NVidia for GPU donation and Capes for financial support.

References

[1] N. Martinel, C. Piciarelli, C. Micheloni, and G. L. Foresti, "On filter banks of texture features for mobile food classification," in *Proceedings of the 9th International Conference on Distributed Smart Cameras*, ser. ICDSC '15. New York, NY, USA: ACM, 2015, pp. 14–19.

[2] F. Ragusa, V. Tomaselli, A. Furnari, S. Battiato, and G. M. Farinella, "Food vs non-food classification," in *Proceedings of the 2Nd International Workshop on Multimedia Assisted Dietary Management*, ser. MADiMa '16. New York, NY, USA: ACM, 2016, pp. 77–81.

[3] P. Pouladzadeh and S. Shirmohammadi, "Mobile multi-food recognition using deep learning," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 3s, pp. 36:1–36:21, Aug. 2017.

[4] M. R. A. Hemad Zareiforoush, Saeid Minaei and A. Banakar, "Qualitative classification of milled rice grains using computer vision and metaheuristic techniques," *J Food Sci Technol*, vol. 53(1), pp. 118–131, 2016.

[5] C. Cusano, P. Napoletano, and R. Schettini, "Combining local binary patterns and local color contrast for texture classification under varying illumination," *J. Opt. Soc. Am. A*, vol. 31, no. 7, pp. 1453–1461, Jul 2014.

[6] T. T. Negri, F. Zhou, Z. Obradovic, and A. Gonzaga, "A robust descriptor for color texture classification under varying illumination," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, vol. 4, 2017, pp. 378–388.

[7] ——, "Extended color local mapped pattern for color texture classification under varying illumination," *Journal of Electronic Imaging*, vol. 27, pp. 27 – 27 – 12, 2018.

[8] R. M. Anwer, F. S. Khan, J. van de Weijer, M. Molinier, and J. Laaksonen, "Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 138, pp. 74 – 85, 2018.

[9] T. T. Negri, R. T. Vieira, and A. Gonzaga, "Color texture classification by using opponent color and local mapped pattern," in *Proceedings of the XIII Workshop of Computer Vision*, 2017, pp. 1–6.

[10] T. Maenpaa and M. Pietikainen, "Classification with color and texture: jointly or separately?" *Pattern Recognition*, vol. 37, no. 8, pp. 1629–1640, 2004.

TABLE III

B1 - CLASSIFICATION ACCURACY (%) OBTAINED BY THE MODELS TRAINED WITH IMAGES ACQUIRED AT LIGHT DIRECTION OF 24 DEGREES.

Models	Direction of the test samples								Avg. Accuracy
	$\theta = 30$	$\theta = 36$	$\theta = 42$	$\theta = 48$	$\theta = 54$	$\theta = 60$	$\theta = 66$	$\theta = 90$	
Alexnet	81.80	53.86	35.11	27.11	22.79	23.80	27.48	33.45	**38.17**
Resnet18	86.48	57.72	34.46	23.80	20.31	21.13	23.89	33.91	37.12
Resnet34	82.81	51.10	32.16	23.43	19.57	19.48	21.87	25.00	34.42
Resnet50	77.38	44.94	26.93	20.40	17.00	17.27	20.58	22.33	30.85
ECLMP	95.40	84.19	60.11	43.38	32.53	24.44	22.24	22.05	48.04
CILMP	93.38	77.94	58.45	46.32	33.63	25.18	24.63	23.71	47.90
OCLMP	94.85	89.15	78.12	67.64	53.86	42.46	36.21	24.44	60.84
OCLBP	94.30	88.23	81.06	75.91	63.78	55.51	50.00	36.58	**68.17**

TABLE IV

B2 - CLASSIFICATION ACCURACY (%) OBTAINED BY THE MODELS TRAINED USING IMAGES ACQUIRED UNDER ALL THE LIGHT DIRECTIONS.

Models	Direction of the test samples									Avg. Accuracy
	$\theta = 24$	$\theta = 30$	$\theta = 36$	$\theta = 42$	$\theta = 48$	$\theta = 54$	$\theta = 60$	$\theta = 66$	$\theta = 90$	
Alexnet	49.26	74.08	90.25	93.75	94.11	93.56	94.85	95.22	87.68	85.86
Resnet 18	59.74	82.72	92.27	96.32	95.77	96.50	97.61	96.50	90.80	**89.80**
Resnet 34	54.77	80.33	93.75	95.58	97.24	96.87	96.50	95.95	89.52	88.94
Resnet 50	49.08	78.30	92.46	96.50	96.69	96.69	95.95	95.58	88.78	87.78
ECLMP	98.53	98.35	98.53	98.90	98.90	99.45	98.90	98.35	98.53	**98.72**
CILMP	95.77	96.69	97.79	98.71	99.08	98.53	98.16	97.98	97.24	97.77
OCLMP	97.61	98.53	98.90	98.71	98.90	98.16	97.98	97.79	97.79	98.26
OCLBP	95.22	95.77	96.88	96.88	96.69	96.51	96.51	95.77	97.06	96.37

TABLE V

C1 - CLASSIFICATION ACCURACY (%) OBTAINED BY THE MODELS TRAINED WITH IMAGES ACQUIRED AT 4000K (D40)

Models	Temperature of the test samples											Avg. Accuracy
	D45	D50	D55	D60	D65	D70	D75	D80	D85	D90	D95	
Alexnet	71.59	31.89	16.54	12.86	9.46	9.28	8.91	8.45	7.44	6.61	6.25	17.20
Resnet 18	67.00	30.33	14.06	7.53	5.33	5.79	5.79	5.88	5.97	5.88	5.97	14.50
Resnet 34	68.56	30.69	15.5	12.04	10.20	9.19	8.23	8.54	8.45	8.08	8.08	17.05
Resnet 50	69.66	31.98	17.09	12.77	10.56	10.20	9.83	9.37	8.45	6.61	4.87	**17.39**
ECLMP	99.08	98.34	97.4	97.05	96.69	95.58	95.58	95.22	95.58	95.40	95.58	96.50
CILMP	98.16	98.89	97.05	97.05	96.87	96.87	97.24	96.87	96.32	95.95	95.77	**97.00**
OCLMP	97.05	90.80	72.61	54.04	42.64	33.08	23.34	19.66	15.99	15.25	13.78	43.48
OCLBP	90.62	79.22	58.27	42.27	32.53	26.65	23.89	20.03	18.19	17.09	17.09	38.71

TABLE VI

C2 - CLASSIFICATION ACCURACY (%) OBTAINED BY THE MODELS TRAINED WITH ALL THE TWELVE TEMPERATURES

Models	Temperature of the test samples												Avg. Accuracy
	D40	D45	D50	D55	D60	D65	D70	D75	D80	D85	D90	D95	
Alexnet	70.22	80.33	83.45	87.31	88.23	86.39	90.62	91.17	91.54	89.15	84.55	80.33	85.27
Resnet 18	91.72	94.11	95.95	96.50	96.87	95.95	96.87	97.24	96.50	95.95	96.69	96.32	**95.88**
Resnet 34	90.07	91.36	94.11	94.30	96.50	95.77	96.32	96.87	95.77	95.95	95.95	95.58	94.87
Resnet 50	89.52	91.17	93.75	95.95	96.32	95.58	96.50	97.79	97.79	97.24	97.61	97.79	95.58
ECLMP	98.90	98.90	99.08	99.26	99.26	99.26	99.45	99.45	99.08	99.45	99.08	99.08	**99.19**
CILMP	98.71	98.90	98.90	98.53	98.72	98.16	98.35	98.53	98.35	98.71	98.35	98.35	98.55
OCLMP	98.90	99.08	99.08	98.90	98.53	98.53	99.08	98.35	98.71	99.08	90.71	98.16	98.09
OCLBP	93.20	95.04	95.40	94.85	94.85	94.12	94.67	94.30	93.57	94.49	96.51	96.88	94.82

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[13] C. Cusano, P. Napoletano, and R. Schettini, "Evaluating color texture descriptors under large variations of controlled lighting conditions," *J. Opt. Soc. Am. A*, vol. 33, no. 1, pp. 17–30, Jan 2016.