Convolutional Neural Network Architectures for the Automated Diagnosis of Celiac Disease

G. Wimmer[1], S. Hegenbart[1], A. Vecsei[2], and A. Uhl[1]

[1] University of Salzburg, Department of Computer Sciences
[2] St. Anna Children's Hospital, Department Pediatrics

Abstract In this work, convolutional neural networks (CNNs) are applied for the computer assisted diagnosis of celiac disease based on endoscopic images of the duodenum. To evaluate which network configurations are best suited for the classification of celiac disease, several different CNN networks were trained using different numbers of layers and filters and different filter dimensions. The results of the CNNs are compared with the results of popular general purpose image representations such as Improved Fisher Vectors and LBP-based methods as well as a feature representations especially designed for the classification of celiac disease. We will show that the deeper CNN architectures outperform these comparison approaches and that combining CNNs with linear support vector machines furtherly improves the classification rates for about 3-7% leading to distinctly better results (up to 97%) than those of the comparison methods.

Keywords: CNN, celiac disease, endoscopy, deep learning

1 Introduction

Convolutional neural networks (CNN) are gaining more and more interest in computer vision. The increase in computational power based on GPUs has led to more sophisticated and deeper architectures which have proven in various challenges to be the state-of-the art in image classification. Generally thousands or millions of images are used and required as data corpus to achieve well generalizing deep architectures. In endoscopic image classification however the available amount of data usable as training corpus is often much more limited to a few hundreds or thousands of images or even less. Another difference to datasets such as used in ILSVRC or Places is however that image classification problems in medical scenarios are often reduced to a few categories instead of thousands in the former. Consequently, deep architectures designed for recognizing images from thousands of categories could be too complex for the classification of celiac disease.

CNNs are already widely used for the computer aided diagnosis in medical scenarios [10], however not so in the computer aided diagnosis using endoscopic imagery. We found only three publications in this area, 2 about the classification of digestive organs using wireless capsule endoscopy images [21,19] and one about lesion detection [20] in endoscopic images. Since the classification of celiac disease can be considered as a texture classification problem and CNNs are state-of-the-art in texture recognition, CNNs are promising image representations for the automated classification of celiac disease.

In this experimental study we apply CNNs for the classification of celiac disease using a experimental setup especially adapted for endoscopic imagery and we try to answer the following open questions:

1. Are deep-architectures suited to classify celiac disease or are simpler and more shallow architectures more suited in such a scenario because of the low amount of training data and categories
2. What are the best network configurations like e.g. the number or filters and their dimensions
3. How well do CNNs perform compared to other state-of-the-art approaches
4. Are linear support vector machines (SVMs) able to furtherly improve the results when applied on the activations of the nets.

2 Celiac Disease

Celiac disease is a complex autoimmune disorder in genetically predisposed individuals of all age groups after introduction of gluten containing food. The gastrointestinal manifestations invariably comprise an inflammatory reaction within the mucosa of the small intestine caused by a dysregulated immune response triggered by ingested gluten protein. During the course of the disease, hyperplasia of the enteric crypts occurs and the mucosa eventually looses its absorptive villi thus leading to a diminished ability to absorb nutrients. [5] state that more than 2 million people in the United States, this is about one in 133, have the disease. People with untreated celiac disease are at risk for developing various complications like osteoporosis, infertility and other autoimmune diseases including type 1 diabetes, autoimmune thyroid disease and autoimmune liver disease. So an early diagnosis is of highest importance.

Endoscopy with biopsy is currently considered the gold standard for the diagnosis of celiac disease. Computer-assisted systems for the diagnosis of CD have potential to improve the whole diagnostic work-up, by saving costs, time and manpower and at the same time increase the safety of the procedure. A motivation for such a system is furthermore given as the inter-observer variability is reported to be high [1,12]. A survey on computer aided decision support for the diagnosis of celiac disease can be found in [9].

Besides standard upper endoscopy, several new endoscopic approaches for diagnosing CD have been evaluated and found their way into clinical practice [2]. The most notable techniques include the modified immersion technique (MIT [7]) under traditional white-light illumination (denoted as WL_{MIT}), as well as MIT under narrow band imaging [3,17] (denoted as NBI_{MIT}). These specialized endoscopic techniques were specifically designed for improving the visual confirmation of CD during endoscopy.

In this work we differentiate between healthy mucosa and mucosa affected by celiac disease using images gathered by NBI_{MIT} as well as WL_{MIT} endoscopy. Examples of the two classes for both endoscopy types are shown in Figure 1. In [6] it was shown that using NBI_{MIT} or WL_{MIT} as imaging modality has a significant impact on the underlying feature distribution of general purpose image representations. However, it was also shown that systems trained on images from both modalities generalize well without requiring additional domain adaption techniques and that combining both modalities
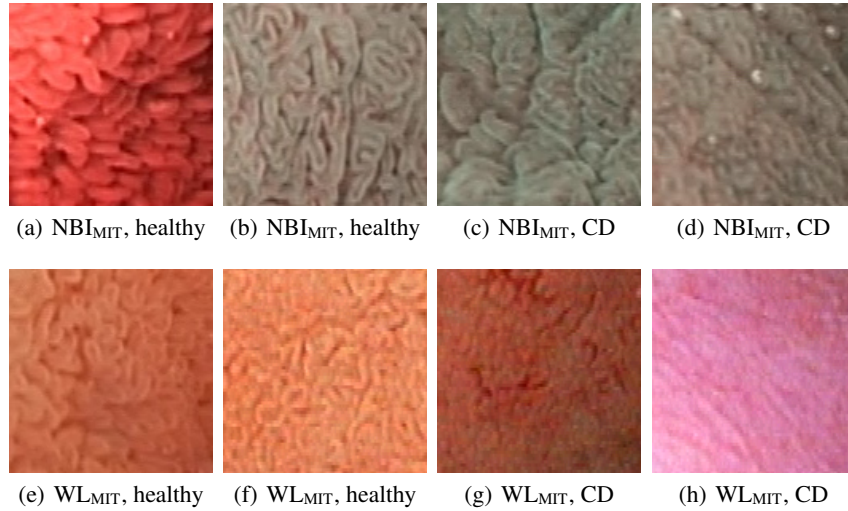
(a) NBI$_{MIT}$, healthy   (b) NBI$_{MIT}$, healthy   (c) NBI$_{MIT}$, CD   (d) NBI$_{MIT}$, CD

(e) WL$_{MIT}$, healthy   (f) WL$_{MIT}$, healthy   (g) WL$_{MIT}$, CD   (h) WL$_{MIT}$, CD

Figure 1. Example images for the two classes healthy and celiac disease (CD) using NBI$_{MIT}$ as well as WL$_{MIT}$ endoscopy

improves the accuracies in case of an insufficient amount of data for training (as is probably the case for CNNs).

3 CNN Architectures

All our networks share the same basic principal architecture. They consist of a variable number of convolutional blocks (CONV) using rectified linear units (RELU) for non-linearity, local response normalization (LRN) [11] and max-pooling (POOL), two fully connected blocks (FC) using RELU and dropout and a last fully connected block acting as soft-max classifier: $[CONV, RELU, LRN, POOL]^n \rightarrow [FC, RELU, DROPOUT]^2 \rightarrow [FC, SOFTMAXLOSS]$. We only vary the number of convolutional blocks, the filter dimensions and the number of filters. To provide a systematic analysis, we trained networks with $n = 1, 2, 3$ and 4 convolutional blocks using different filter dimensions and different numbers of filters in each layer. We follow the general approach of employing large filter dimensions in lower layers and subsequently smaller filters in higher layers.

A high number of filters per layer allows the training process to adapt to highly abstract features. However, it is unclear in the context of celiac disease and endoscopic imagery in general if such abstract features are visible or even useful for prediction. Consequently, we analyze the impact of the number of filters per layer by training multiple nets of the same architecture with varying numbers of filters. We generally rely on the concept of increasing the number of filters from the lower to the higher layers by a factor of two per layer.

All our models are initialized and trained using the same set of techniques. The coefficients of the nets are randomly initialized based on He et al. [8] and the bias terms are initialized as 0. All architectures rely on using max-pooling with a windows size

of three and stride two. Stochastic gradient descent (SGD) with weight decay ($\lambda = 0.0005$) and momentum ($\mu = 0.9$) is used for the training of the models. Regularization is achieved using drop-out ($p = 0.5$) during training. Training is performed on batches of 128 images each, which are for each iteration randomly chosen from the training data and subsequently augmented (see Section 4.1). The learning rate is initialized at 0.01 and four times divided by three whenever the training-loss stopped improving with the current learning rate. For this, each 250th iteration we compute the average loss of the previous 250 iterations. If the currently computed average loss is greater than 0.99 times the previously computed average loss and if the current learning rate is in use for at least 1000 iterations, then the learning rate is divided by three. Due to the differing number of parameters among the architectures, optimization is continued until the training-loss shows no improvement over 2500 iterations but at least until the learning rate has been reduced the fourth time. The model of the iteration achieving the lowest training-loss is then used for validation.

Our learning rate configurations and break off condition are especially adapted on our celiac disease image data to achieve high results without needing too much time for training (the nets were trained for ≈ 10000 iterations in average). Since we train 36 different nets (4 (different numbers of convolutional blocks) $\times$ 3 (different filter sizes) $\times$ 3 (different filter numbers)) on 10 different training splits (see Section 4.1), we had to choose such configurations that enable a limited time of training per network.

3.1 Very-Shallow Networks

We start off with a very uncommon variation of CNNs using only one single convolutional block. By analyzing different architectures growing from very shallow to deep we hope to gain some insight on the problem. Although this sort of architecture is quite uncommon and might not fit into the general CNN schemes, the lower abstraction of features in endoscopic images and the small number of categories (two) make it necessary to start with such shallow architectures. The Very-Shallow networks (see Table 1) are trained with $N = 10, 48$ and 96 filters to analyze the impact of the number of filters on the results.

Filter size	$CONV_1$	FC_1	FC_2	FC_3
Large	$11 \times 11 \times N$	512	512	2
	st. 3, pad 0	drop-out	drop-out	soft-max
Medium	$7 \times 7 \times N$	512	512	2
	st. 3, pad 0	drop-out	drop-out	soft-max
Small	$5 \times 5 \times N$	512	512	2
	st. 2, pad 0	drop-out	drop-out	soft-max

Table 1. Architecture of the Very-Shallow networks. The first row in a convolutional block (CONV) specifies the receptive field size of the convolutional filters and their number (N). The second row indicates the stride (st.) and padding (pad). Furtherly we indicate the dimensionality of the fully connected (FC) blocks.

3.2 Shallow Networks

The next generation of architectures is based on the Very-Shallow networks but the number of convolutional blocks is increased to two. Like in the previous and also in the following deeper network architectures, the network is trained with different numbers of filters ($N = 10, 48$ and 96 filters in the first convolutional layer). The network architecture of the Shallow nets is shown in Table 2.

Filter size	CONV$_1$	CONV$_2$	FC$_1$	FC$_2$	FC$_3$
Large	$11 \times 11 \times N$	$7 \times 7 \times 2N$	512	512	2
	st. 3, pad 0	st. 3, pad 0	drop-out	drop-out	soft-max
Medium	$7 \times 7 \times N$	$5 \times 5 \times 2N$	512	512	2
	st. 4, pad 0	st. 2, pad 0	drop-out	drop-out	soft-max
Small	$5 \times 5 \times N$	$3 \times 3 \times 2N$	512	512	2
	st. 3, pad 0	st. 2, pad 0	drop-out	drop-out	soft-max

Table 2. Architecture of the Shallow networks.

3.3 Deep Networks

The third generation of nets use 3 convolutional blocks and can therefore be considered as our first deep architecture. The network architecture of the Deep nets is shown in Table 3.

Filter size	CONV$_1$	CONV$_2$	CONV$_3$	FC$_1$	FC$_2$	FC$_3$
Large	$11 \times 11 \times N$	$7 \times 7 \times m_{2N}^{128}$	$5 \times 5 \times m_{4N}^{256}$	512	512	2
	st. 2, pad 0	st. 1, pad 0	st. 1 pad 0	drop-out	drop-out	soft-max
Medium	$7 \times 7 \times N$	$5 \times 5 \times m_{2N}^{128}$	$3 \times 3 \times m_{4N}^{256}$	512	512	2
	st. 2, pad 0	st. 1, pad 0	st. 1, pad 0	drop-out	drop-out	soft-max
Small	$5 \times 5 \times N$	$3 \times 3 \times m_{2N}^{128}$	$3 \times 3 \times m_{4N}^{256}$	512	512	2
	st. 2, pad 0	st. 1, pad 0	st. 1, pad 0	drop-out	drop-out	soft-max

Table 3. Architecture of the Deep networks, where $m_b^a = \max(a, b)$ and denotes the number of convolutional filters.

3.4 Very-Deep Networks

In our last generation of nets we use 4 convolutional blocks (see Table 4). Although the term Very-Deep is not quite true considering the number of layers of other very-deep architectures, we use the term to easily distinguish between our four basic architectures.

Filter size	CONV$_1$	CONV$_2$	CONV$_3$	CONV$_4$	FC$_1$	FC$_2$	FC$_3$
Large	11x11xN	7x7xm_{2N}^{192}	5x5xm_{4N}^{256}	3x3xm_{8N}^{256}	1024	1024	2
	st. 1, pad 2	st. 1, pad 0	st. 1, pad 0	st. 1, pad 0	drop-out	drop-out	soft-max
Medium	7x7xN	5x5xm_{2N}^{192}	3x3xm_{4N}^{256}	3x3xm_{8N}^{256}	1024	1024	2
	st. 2, pad 0	st. 1, pad 0	st. 1, pad 0	st. 1, pad 0	drop-out	drop-out	soft-max
Small	5x5xN	3x3xm_{2N}^{192}	3x3xm_{4N}^{256}	3x3xm_{8N}^{256}	1024	1024	2
	st. 2, pad 0	st. 1, pad 0	st. 1, pad 0	st. 1, pad 0	drop-out	drop-out	soft-max

Table 4. Architecture of the Very-Deep networks, where $m_b^a = \max(a, b)$.

4 Experimental Setup and Results

4.1 Experimental Setup

Our celiac disease image database consists of 1661 RGB image patches of size 128×128 pixels that are gathered by means of flexible endoscopes using NBI$_{MIT}$ as well as WL$_{MIT}$. The database consists of 1045 images gathered by WL$_{MIT}$ endoscopy (587 healthy images and 458 affected by celiac disease) and 616 images gathered by NBI$_{MIT}$ endoscopy (399 healthy images and 217 affected by celiac disease). So in total 986 image patches show healthy mucosa and the remaining 675 image patches show mucosa affected by celiac disease. The images were captured from 353 patients.

Due to the relatively small amount of data, we perform cross-validation to achieve a stable estimation of the generalization error. We generated 10 (fixed) splits for training and validation (80% training and 20% validation) and took care that images of a single patient are never in training and evaluation sets. All nets are trained using the training portion of our data corpus. The final validation was performed on the left-out part.

The image data is normalized by subtracting the mean image of the training portion. We then linearly scale each image within $[-1, 1]$. Due to the small amount of available data we use data augmentation to increase the number of images for training. Augmentation is applied to the batches of images extracted for training. The augmentation is based on cropping one sub-image (112×112 pixels) from each training image with randomly chosen position. Subsequently, the sub-image is randomly rotated ($0°$, $90°$, $180°$ or $270°$) and randomly either horizontally reflected or not. Validation is performed using a majority voting of five crops from the validation image using the upper left, upper right, lower left, lower right and center part.

In our experiments, we compute the overall classification rate (OCR) for each split and report the mean OCR over all 10 splits with the respective standard deviation.

The CNNs are implemented using the *MatConvNet* framework [18]. Additionally to the CNN soft-max-classifier we employ linear SVMs as provided by the *LIBLINEAR* library [4]. For this, the training and test samples are fed through the CNNs and the output of the second fully connected layer is extracted as feature for further SVM classification. The size of the extracted feature vector per image is 1024×1 in case of the very-deep architectures and 512×1 for the other architectures. Augmentation is also applied for the extraction of features from the nets for further for SVM classification. The augmentation is basically the same as for the training of the nets with only one dif-

ference. The patches of the training images are extracted from the fixed center position instead from random positions (8 patches per image with 4 different rotations, either horizontally flipped or not). The SVM cost factor (C) is found using cross validation on the training data.

Additionally, we combine CNNs, principle component analysis (PCA) and SVMs by applying PCA to the CNN features resulting in 100 principal components which are furtherly classified using SVMs.

We compare the CNNs against three popular general purpose image representations and one feature representations especially developed for the classification of celiac disease. As general purpose image representations we use multi-resolution local binary patterns (LBP [13]) and multi-resolution local ternary patterns (LTP [15]), both with 3 scales, 8 neighbors and uniform patterns. As third general purpose method we employ the improved fisher vectors (IFV [14]) computed from SIFT descriptors on a dense 6×6 pixel grid. The fourth method, further denoted as fractal analysis based method (FRAC [16]), was especially developed for the classification of celiac disease and is based on pre-filtering images using the rotation invariant MR8 filterbank, followed by computing the local fractal dimension (see [16]) of the resulting filter responses and applying the bag-of-visual words (BoW) approach to them. We rely on in-house MAT-LAB implementations for LBP, LTP and FRAC and use the implementation of IFV as provided by *VLFeat*. The comparison methods are classified using SVMs in an analogous manner as for the CNN features.

4.2 Results

The results of our experiments are presented in Table 5. The standard deviations are given in brackets. The best result of each network architecture and classification strategy is given in bold face numbers.

As we can seen in Table 5, the highest CNN results are achieved using the Deep and Very-Deep network architectures combined with large or medium sized filters. Using only 10 filters in the first convolutional layer is insufficient for the classification of celiac disease, but using 48 filters achieves similar results as using 96. The two deeper CNN architectures with large or medium sized filters achieve classification rates of $\approx 90\%$ and hence outperform the comparison methods, whose highest classification rate is 89.5% (LTP). Combining CNNs and SVMs furtherly improves the results for about 3-7%. Additionally applying PCA to the CNN features has only a minimal effect to the results. The best results ($\approx 97\%$) are achieved using SVM classification (with or without PCA) applied to the CNN features of the Very-Deep net with 96 filters of size $11 \times 11 \times 3$ in the first convolutional layer.

5 Conclusion

In this work we showed that deep CNN architectures are very suited for the classification of celiac disease based on endoscopic image data. These CNN networks outperform other state-of-the-art image representation approaches. Simpler and more shallow-architectures cannot compete with the deeper architectures. Using large or medium filter dimensions generally leads to higher results than using smaller filter dimensions.

Very-shallow networks

Nr. of filters \ Size	CNN large	medium	small	CNN & SVM large	medium	small	CNN & SVM & PCA large	medium	small
10	85.8(2.3)	85.9(1.7)	86.0(2.0)	89.5(1.9)	88.6(1.9)	91.1(2.9)	89.1(2.1)	88.7(1.6)	91.0(2.7)
48	**88.0**(1.1)	87.4(1.6)	87.6(1.6)	92.5(2.4)	93.3(3.0)	**94.4**(2.9)	92.7(2.7)	93.3(2.7)	**94.3**(3.1)
96	87.5(1.3)	86.8(1.7)	87.9(2.0)	92.1(2.2)	92.8(2.4)	93.3(2.2)	92.2(2.5)	92.6(2.5)	93.3(2.5)

Shallow networks

Nr. of filters \ Size	CNN large	medium	small	CNN & SVM large	medium	small	CNN & SVM & PCA large	medium	small
10	82.6(4.2)	86.1(2.7)	86.3(1.5)	87.2(2.3)	88.9(1.6)	88.2(2.1)	86.9(2.4)	88.8(1.6)	87.9(2.2)
48	88.5(2.6)	89.9(1.3)	89.1(1.3)	92.1(2.1)	93.3(2.3)	93.2(1.6)	92.0(2.1)	93.3(2.4)	92.9(1.8)
96	88.5(1.4)	**90.0**(1.8)	89.6(1.4)	92.3(2.4)	**94.1**(2.6)	92.9(2.1)	92.4(2.4)	**94.1**(2.4)	93.1(2.2)

Deep networks

Nr. of filters \ Size	CNN large	medium	small	CNN & SVM large	medium	small	CNN & SVM & PCA large	medium	small
10	87.9(1.4)	89.6(1.4)	88.9(1.2)	93.2(2.6)	92.9(2.1)	92.4(2.1)	93.3(2.6)	93.0(2.2)	92.2(2.2)
48	89.8(1.6)	**90.5**(1.6)	89.8(1.3)	**96.7**(3.0)	96.4(2.8)	95.9(2.4)	**96.6**(3.0)	96.4(2.9)	95.7(2.3)
96	89.1(1.7)	89.9(1.3)	89.4(1.7)	96.5(3.2)	96.4(2.7)	95.4(3.3)	96.5(3.2)	**96.6**(2.6)	95.5(3.0)

Very-deep networks

Nr. of filters \ Size	CNN large	medium	small	CNN & SVM large	medium	small	CNN & SVM & PCA large	medium	small
10	88.7(1.4)	88.2(2.2)	88.0(1.6)	94.6(2.6)	93.0(2.2)	91.6(2.2)	94.7(2.7)	93.0(2.3)	91.8(2.3)
48	89.5(1.8)	89.3(2.0)	89.2(1.9)	96.5(3.7)	95.7(3.1)	95.6(2.6)	96.5(3.7)	95.7(3.1)	95.5(2.5)
96	**90.3**(1.7)	89.8(1.6)	89.4(1.3)	**97.0**(3.1)	96.5(2.6)	95.4(3.3)	**97.1**(3.4)	96.5(2.6)	95.3(3.1)

Comparison methods

LBP	LTP	IFV	FRAC
86.4(2.7)	**89.5**(1.8)	84.7(2.8)	80.1(3.9)

Table 5. Results of the CNNs and comparison methods

Applying SVMs on the activations of the nets furtherly improves the results of the CNNs for about 3-7% up to a maximum of $\approx 97\%$. The highest result was achieved using SVM classification, the deepest architecture (Very-Deep), the largest filter dimension and the highest number of filters (96 filters of size $11 \times 11 \times 3$ in the first convolutional layer).

References

1. Biagi, F., Rondonotti, E., Campanella, J., Villa, F., Bianchi, P.I., Klersy, C., Franchis, R.D., Corazza, G.R.: Video capsule endoscopy and histology for small-bowel mucosa evaluation: A comparison performed by blinded observers. Clinical Gastroenterology and Hepatology 4(8), 998–1003 (2006)
2. Chand, N., Mihas, A.A.: Celiac disease: Current concepts in diagnosis and treatment. Journal of Clinical Gastroenterology 40(1), 3–14 (January 2006)
3. Emura, F., Saito, Y., H., I.: Narrow-band imaging optical chromocolonoscopy: advantages and limitations. World J Gastroenterol 14(31), 4867–4872 (2008)
4. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: A library for large linear classification. J. Mach. Learn. Res. 9, 1871–1874 (Jun 2008)
5. Fasano, A., Berti, I., Gerarduzzi, T., Not, T., Colletti, R.B., Drago, S., Elitsur, Y., Green, P.H.R., Guandalini, S., Hill, I.D., Pietzak, M., Ventura, A., Thorpe, M., Kryszak, D.,

Fornaroli, F., Wasserman, S.S., Murray, J.A., Horvath, K.: Prevalence of celiac disease in at-risk and not-at-risk groups in the united states: a large multicenter study. Archives of internal medicine 163, 286–92 (February 2003)

6. Gadermayr, M., Hegenbart, S., Kwitt, R., Uhl, A.: Narrow band imaging versus white-light: What is best for computer-assisted diagnosis of celiac disease? In: Proc. of the 13th IEEE International Symposium on Biomedical Imaging (ISBI'16). pp. 355–359 (Apr 2016)

7. Gasbarrini, A., Ojetti, V., Cuoco, L., Cammarota, G., Migneco, A., Armuzzi, A., Pola, P., Gasbarrini, G.: Lack of endoscopic visualization of intestinal villi with the immersion technique in overt atrophic celiac disease. Gastrointestinal endoscopy 57, 348–351 (2003)

8. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. CoRR (2015)

9. Hegenbart, S., Uhl, A., Vécsei, A.: Survey on computer aided decision support for diagnosis of celiac disease. Computers in Biology and Medicine (65), 348–358 (2015)

10. Jiang, J., Trundle, P., Ren, J.: Medical image analysis with artificial neural networks. Computerized Medical Imaging and Graphics 34(8), 617 – 631 (2010)

11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25, pp. 1097–1105 (2012)

12. Niveloni, S., Fiorini, A., Dezi, R., Pedreira, S., Smecuol, E., Vazquez, H., Cabanne, A., Boerr, L.A., Valero, J., Kogan, Z., Maurino, E., Bai, J.C.: Usefulness of videoduodenoscopy and vital dye staining as indicators of mucosal atrophy of celiac disease: assessment of interobserver agreement. Gastrointestinal Endoscopy 47(3), 223–229 (1998)

13. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recogn 29(1), 51–59 (1996)

14. Perronnin, F., Liu, Y., Sanchez, J., Poirier, H.: Large-scale image retrieval with compressed fisher vectors. In: Proceedings of CVPR'10. pp. 3384–3391 (2010)

15. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Proceedings of AMFG'07. pp. 168–182 (2007)

16. Uhl, A., Vécsei, A., Wimmer, G.: Fractal analysis for the viewpoint invariant classification of celiac disease. In: Proceedings of the 7th International Symposium on Image and Signal Processing (ISPA 2011). pp. 727 –732. Dubrovnik, Croatia (Sep 2011)

17. Valitutti, F., Oliva, S., Iorfida, D., Aloi, M., Gatti, S., Trovato, C.M., Montuori, M., Tiberti, A., Cucchiara, S., Di Nardo, G.: Narrow band imaging combined with water immersion technique in the diagnosis of celiac disease. Dig and Liver Dis 46(12), 1099–1102 (2014)

18. Vedaldi, A., Lenc, K.: Matconvnet – convolutional neural networks for matlab. In: Proceeding of the ACM Int. Conf. on Multimedia. pp. 689–692 (2015)

19. s. Yu, J., Chen, J., Xiang, Z.Q., Zou, Y.X.: A hybrid convolutional neural networks with extreme learning machine for wce image classification. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), 2015. pp. 1822–1827 (Dec 2015)

20. Zhu, R., Zhang, R., Xue, D.: Lesion detection of endoscopy images based on convolutional neural network features. In: 8th International Congress on Image and Signal Processing (CISP). pp. 372–376 (Oct 2015)

21. Zou, Y., Li, L., Wang, Y., Yu, J., Li, Y., Deng, W.J.: Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network. In: IEEE International Conference on Digital Signal Processing (DSP), 2015. pp. 1274–1278 (July 2015)