CNN Transfer Learning for the Automated Diagnosis of Celiac Disease

G. Wimmer*, A. Vécsei† and A. Uhl*
* University of Salzburg, Department of Computer Sciences, Salzburg, Austria
e-mail: gwimmer@cosy.sbg.ac.at, uhl@cosy.sbg.ac.at
† St. Anna Children's Hospital, Department Pediatrics,
Medical University, Vienna, Austria

***Abstract*—In this work, four well known convolutional neural networks (CNNs) that were pretrained on the ImageNet database are applied for the computer assisted diagnosis of celiac disease based on endoscopic images of the duodenum. The images are classified using three different transfer learning strategies and a experimental setup specifically adapted for the classification of endoscopic imagery. The CNNs are either used as fixed feature extractors without any fine-tuning to our endoscopic celiac disease image database or they are fine-tuned by training either all layers of the CNN or by fine-tuning only the fully connected layers. Classification is performed by the CNN SoftMax classifier as well as linear support vector machines. The CNN results are compared with the results of four state-of-the-art image representations. We will show that fine-tuning all the layers of the nets achieves the best results and outperforms the comparison approaches.**

***Index Terms*—Convolutional neural networks, Transfer learning, Celiac Disease, Automated diagnois, Endoscopy.**

I. 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. Such datasets are often constructed using crowd-sourcing and are often based on queries of image-search engines such as Google.

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. Consequently it is hard to achieve generalization with existing deep architectures on such endoscopic data. 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.

CNNs have already been widely used for the computer aided diagnosis in medical scenarios [1], however not for endoscopic imagery. We found only three publications in this area, 2 about the classification of digestive organs using wireless capsule endoscopy images [2], [3] and one about lesion detection [4] in endoscopic images. Since CNNs are now one of the dominant approach for feature extraction from texture data and the automated diagnosis of celiac disease is usually considered as a texture classification problem, CNNs could be promising image representations for the classification of celiac disease.

However, because of the limited amount of endoscopic celiac disease images (1661 images), CNNs that are trained on those images would almost certainly be overfitted to the training data. One way to avoid overfitting on small databases is to use CNNs that were pre-trained on huge databases, a process known as transfer learning. In our experimental study we use four well known CNNs, the AlexNet [5] and three VGG nets [6]. All these nets were trained on the ImageNet ILSVRC challenge data (http://www.image-net.org/challenges/LSVRC/).

It has been observed that the features of earlier layers of a CNN contain more generic features (e.g. edge detectors or color blob detectors) that should be useful to many tasks, but later layers of the CNN become progressively more specific to the details of the classes contained in the original dataset [7]. Thus the CNN parameters from earlier layers may require less or entirely no fine-tuning to the target data. In [8], features were extracted from pre-trained CNNs without any fine-tuning to the target database. The features were extracted from the first fully connected layer of the CNNs and were classified using linear support vector machines (SVMs). This approach outperformed other state-of-the-art image representations on several different image databases.

Based on these findings of [7] and [8], we perform three transfer learning scenarios:

1) We use CNNs as fixed feature extractors without any fine-tuning to our dataset. Features are extracted from the first fully connected layer and classification is done using a linear SVM.
2) We fine-tune the fully connected layers (the layers specific to the classes contained in the original dataset) and keep the previous layers (the convolutional layers) fixed. Classification is done using the CNN-built in SoftMax classifier as well as a SVM using features of the last fully connected layer.
3) We fine tune all layers of the CNN. For the Classification we use the CNN SoftMax classifier and SVMs using the features of the first and the last fully connected layer.

978-1-4673-8910-5/16/$31.00 ©2016 IEEE

In this way we want to find out which transfer learning strategy is most suited for the classification of our celiac disease database. That means we want to find out how much training should be performed for the classification of celiac disease. Is it best to perform no training to avoid overfitting, or is it better to only train the more data specific layers (the fully connected layers) or is it best to train all layers. We also want to find out which nets are best suited to classify celiac disease and how well does our CNN transfer learning approaches perform compared to state-of-the-art image representations.

II. Celiac Disease

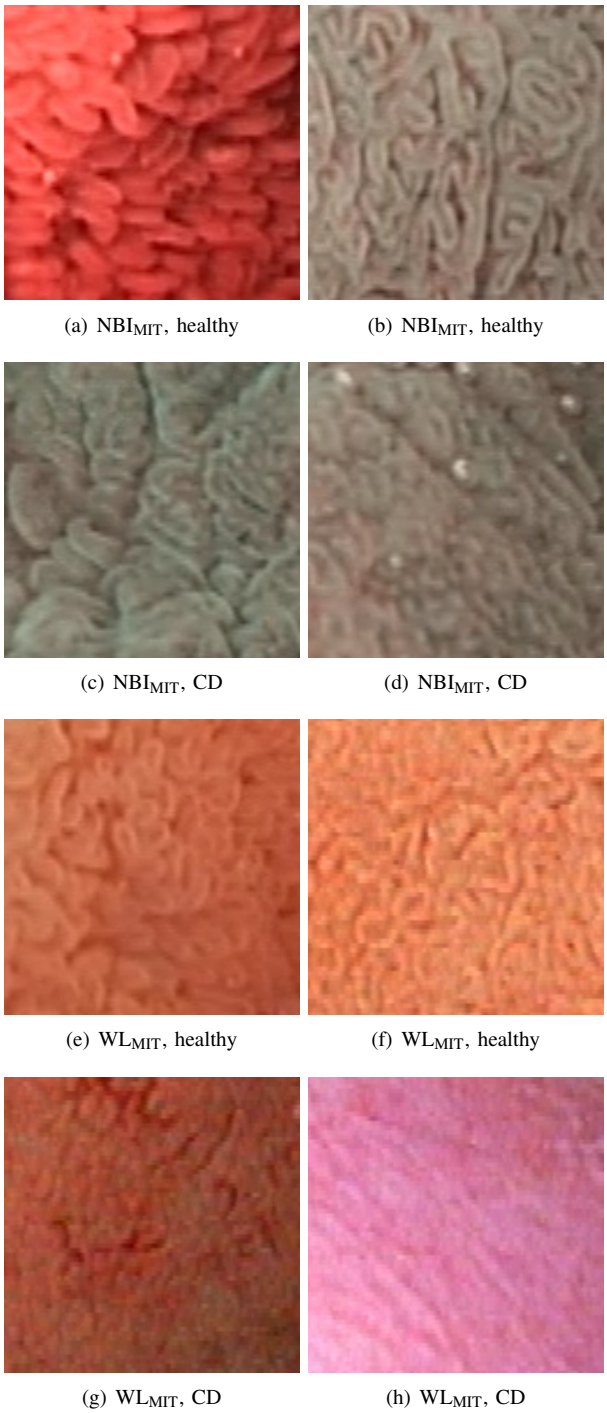
Celiac disease (CD) is a multisystemic immune-mediated disease, which is associated with considerable morbidity and mortality [9]. In untreated or inappropriately treated CD the inflammation caused by the dysregulated immune response can disrupt the intestinal mucosa thus leading to a total atrophy of the villi (finger-like projections of the mucosa) which causes a diminished ability to absorb nutrients. After embarking on a strict gluten-free diet, which is the CD treatment modality of first choice, the inflammation gradually subsides allowing for mucosal healing. To avoid the most severe complications of CD, an early diagnosis for commencing a strict gluten-free diet is of vital importance.

[10] 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.

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 [11], [12].

Besides standard upper endoscopy, several new endoscopic approaches for diagnosing CD have been evaluated and found their way into clinical practice [13]. The most notable techniques include the modified immersion technique (MIT [14]) under traditional white-light illumination (denoted as WL_{MIT}), as well as MIT under narrow band imaging [15], [16] (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 [17] 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 improves the accuracies in case of an insufficient amount of data for training (as is probably the case for CNNs).



(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

III. CNN Transfer Learning

This section gives the implementation details for applying transfer learning to our celiac disease image database.

A. AlexNet and the VGG nets

The first work that popularized CNNs in Computer Vision was about the AlexNet [5]. AlexNet consists of 5 convolutional layers and three fully connected layers with a final SoftMax classifier (see Figure 2).

Also the VGG nets [6] consist of 5 convolutional layers and three fully connected layers with a final SoftMax classifier. There are three versions of these nets. The VGG-f architecture is the fastest of the three nets and is similar to the AlexNet, however with a dense connectivity between convolutionaly layers instead of sparse connections as in case of the AlexNet. The medium fast VGG-m net and the slow VGG-s net have decreased strides and smaller receptive fields which was shown to be beneficial on the ILSVRC dataset but also slows downs the nets.

B. Training of the nets

Our 4 CNNs are initialized and trained using the same set of techniques. As initialization for the convolutional layers we use the parameters that were learned on the ImageNet ILSVRC challenge data. Since the fully connected layers are more specific to the details of the classes contained in the ILSVRC challenge data, we randomly initialize the coefficients of these layers based on He et al. [18] instead of the parameters that were learned on the ImageNet ILSVRC challenge data. The bias terms of the fully connected layers are initialized as 0. The size of the last fully connected layer is adapted to our 2-class classification scheme which means that the size of the convolutional filters are changed from $1 \times 1 \times 4096 \times 1000$ to $1 \times 1 \times 4096 \times 2$. The last fully connected block is acting as soft-max classifier and computes the training loss (log-loss). Training is performed on batches of 128 images each, which are for each iteration randomly chosen from the training data and subsequently augmented. Stochastic gradient descent (SGD) with weight decay ($\lambda = 0.0005$) and momentum ($\mu = 0.9$) is used for the training of the models.

As already mentioned in the introduction, we follow two learning strategies. In one strategy all the layers of the nets are fine-tuned and in the other strategy only the fully connected layers are fine-tuned by setting the learning rate to 0 for the convolutional layers. In case of the fully fine-tuned nets, training is performed for 5000 iterations on our celiac disease image database. We begin training with a learning rate of 0.01 and every 250 iterations the learning rate is decreased by a factor of $f \approx 1.27$ down to a learning rate of 0.0001 for the last 250 iterations.

For the second learning strategy, where only the fully connected layers are fine-tuned, we have seen in experiments that more iterations are required to train the nets We start with 5000 iterations with learning rate 0.01 followed by 10 0000 iterations with stepwise decreasing learning rates (similar to the fully fine-tuned nets, we begin with a learning rate of 0.01 and each 500 iterations the learning rate is decreased by a factor of $f \approx 1.27$ down to a learning rate of 0.0001 for the last 500 iterations).

On the left side of Figure 3 we see the convolutional kernels of the first convolutional layer of the four nets learned on the ImageNet ILSVRC challenge data and on the right side we see the kernels after fine-tuning on our celiac disease database. We can observe that fine-tuning only slightly changed the filter kernels. However, thats exactly what we want to have for the earlier layers since we could not achieve such nice and smooth filters by exclusively train them on our small database. Only in case of the AlexNet some of the filter kernels changed completely. Some of the filters of the AlexNet changed their colors (e.g. the filter on the right side of the third row in Figure 3(a,b)) some others loose their entire structure and turned into simple averaging filters (e.g. the left filter in the first row in Figure 3(a,b)). In case of the three VGG nets, the filter kernels undergo only very minor changes.

C. Feature extraction for SVM classification

Additionally to the CNN SoftMax classifier, the images are classified by SVMs using features extracted from the CNNs. For this, the training and test samples are fed through the CNNs and the input of the first fully connected layer (further denoted as 'fc1') respectively the last fully connected layer (further denoted as 'fcL') is extracted as feature for further SVM classification. In case of an extraction of the input of the first fully connected layer, the size of the extracted features per image is $6 \times 6 \times 256$ (resulting in a feature vector of length 9216 per image) and in case of an extraction of the input of the last fully connected layer the size of the extracted features is 4096×1.

IV. Comparison Methods

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 apply a multiscale block binary patterns (MB-LBP) operator [19] with three different block sizes (3,9,15) and uniform patterns. As second general purpose method we employ the dual-tree complex wavelet transform (DT-CWT [20]) using 4 decomposition levels and we extract the means and standard deviations of the subbands as features. As third general purpose method we employ the improved fisher vectors (IFV [21]) computed from SIFT descriptors on a dense 6×6 pixel grid. The fourth method, further denoted as fractal analysis based method (FRAC [22]), 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 [22]) of the resulting filter responses and applying the bag-of-visual words (BoW) approach to them. We rely on in-house MATLAB implementations for

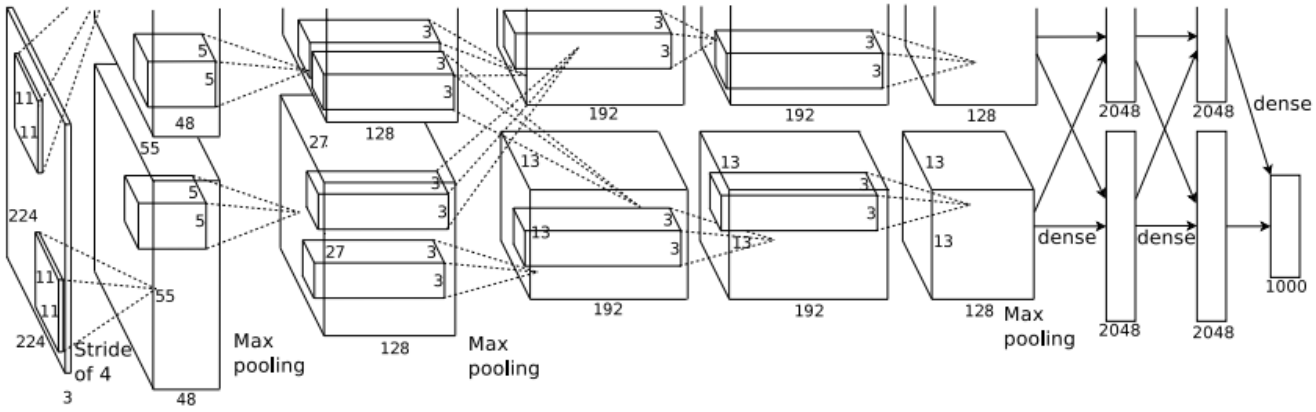


Figure 2. Illustration of the of the AlexNet architecture [5]

MB-LBP, DT-CWT and FRAC and use the implementations of IFV as provided by *VLFeat*.

V. Experimental Setup and Results

A. Experimental Setup

The 1661 RGB image patches of size 128×128 pixels are gathered by means of flexible endoscopes using NBI_{MIT} as well as WL_{MIT}. The celiac disease image 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.

Since our used nets require input image sizes of $227 \times 227 \times 3$ (AlexNet) respectively $224 \times 224 \times 3$ (VGG nets), the image patches are bicubic upscaled to a size of $256 \times 256 \times 3$. The 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 and validation. Augmentation is applied to the batches of images extracted for training. The augmentation is based on cropping one sub-images (227×227 respectively 224×224 pixels) from each image with randomly chosen position. Subsequently, the sub-image is randomly rotated (0°, 90°, 180° or 270°) and randomly either flipped or not flipped around the horizontal axis. 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.

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 difference. 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 and either horizontally flipped or not).

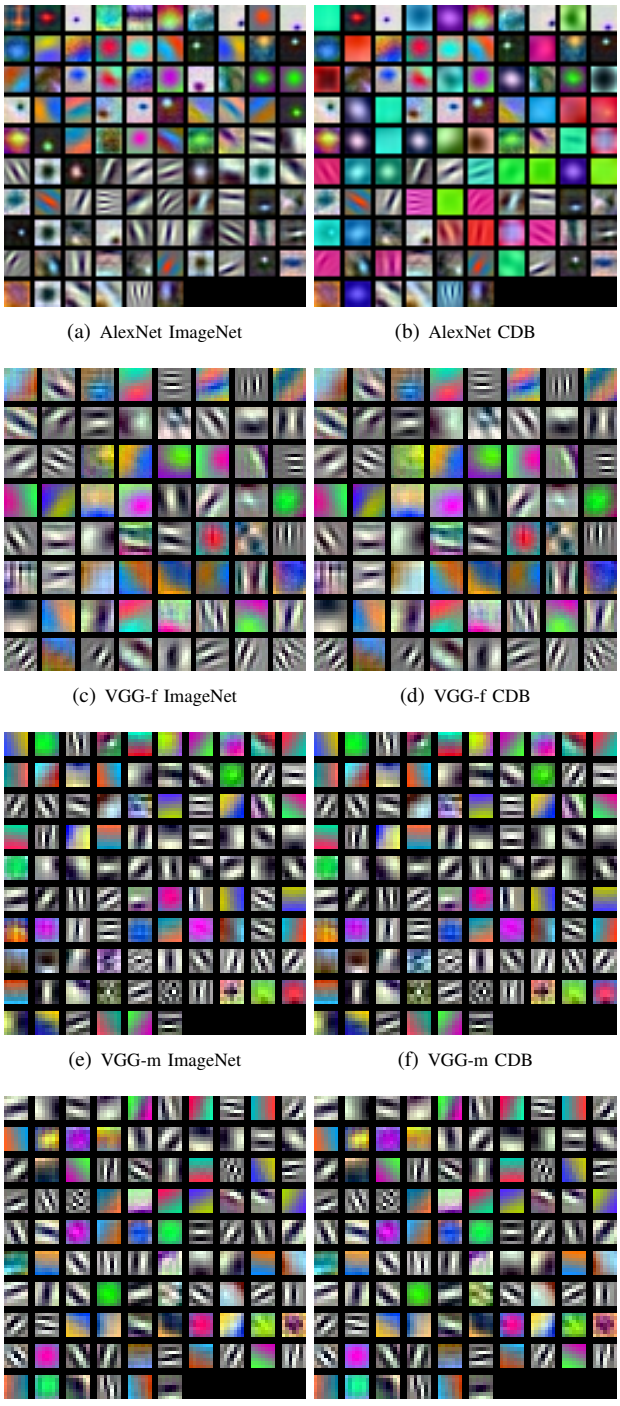
Due to the relatively small amount of data, we perform 5-fold cross-validation to achieve a stable estimation of the generalization error. For each of the 5 folds we took care that images of a single patient are never in training and evaluation sets. All nets are trained using the training portion of our data corpus. The final validation was performed on the left-out part. That means beginning with a net, we train five different nets, one for each of the 5 folds (except for the case where we use the CNNs as fixed feature extractor, where no training is performed). In our experiments, we compute the overall classification rate (OCR) for each fold and report the mean OCR over all 5 folds with the respective standard deviation.

The CNNs are implemented using the *MatConvNet* framework [23]. Additionally to the CNN soft-max-classifier we employ SVMs as provided by the *LIBLINEAR* library [24]. The SVM cost factor (C) is found using cross validation on the training data.

The comparison methods are also classified using SVMs in an analogous manner (5-fold cross validation) as for the CNN features.

B. Experimental Results

The results of our transfer learning experiments are presented in Table I. The standard deviations are given in brackets. For each CNN, the best result over the different learning and classification strategies is given in bold face numbers. The 3 columns titled as 'none', 'fully connected layers' and 'all layers' in Table I indicate which of our three transfer learning strategies are used. 'none' indicates that no learning is performed, 'fully connected layers' that only the fully connected layers are trained (further denoted as partly fine-tuning) and 'all layers' that each layer of the network is trained (further denoted as fully fine-tuning). 'SVM fc1' indicates that the CNN features for the SVM classification are extracted from the first fully connected layer and 'SVM fcL' that the



(a) AlexNet ImageNet (b) AlexNet CDB

(c) VGG-f ImageNet (d) VGG-f CDB

(e) VGG-m ImageNet (f) VGG-m CDB

(g) VGG-s ImageNet (h) VGG-s CDB

Figure 3. Convolutional kernels of the first convolutional layer learned on the ImageNet database or learned on our celiac disease database (CDB) using the parameters learned from the ImageNet ddatabase as initialization

CNN features are extracted from the last fully connected layer. 'SoftMax' means that the CNN SoftMax classifier is used for classification.

As we can see in Table I, the clearly best CNN results are achieved using the fully fine-tuned nets. For these fully fine-tuned CNNs it turrned out to be irrelevant to the outcomes if

CNNs	Training layers					
	none	fully connected layers		all layers		
	SVM fc1	SVM fcL	SoftMax	SVM fc1	SVM fcL	SoftMax
AlexNet	81.1(1.7)	80.7(1.4)	78.5(2.8)	**86.9**(2.1)	86.2(1.9)	86.6 (0.9)
VGG-f	84.9(1.5)	85.7(1.3)	84.4(1.1)	90.4(1.9)	90.4(0.7)	**90.5**(0.7)
VGG-m	85.0(1.8)	85.4(0.6)	83.5(1.3)	89.8(1.7)	**90.0**(1.9)	88.4(1.5)
VGG-s	84.3(2.9)	84.8(3.4)	82.5(1.5)	**89.2**(2.1)	88.6(1.8)	**89.2**(1.7)
MB-LBP	87.0 (2.4)					
DT-CWT	77.3 (1.0)					
IFV	84.4 (2.9)					
Frac	79.2 (2.3)					

Table I
RESULTS OF THE CNNS AND COMPARISON METHODS IN %.

the classification is done using SoftMax or SVM classifiers. It also does not really make a difference if the features for the SVM classification are extracted from the first or the last convolutional layer. In case of the partly fine-tuned CNNs, the SVM results are higher than the SoftMax results. The SVM results for the CNNs without any fine-tuning are similar to the SVM results of the partly fine-tuned CNNs.

When we compare the results of the four CNNs, then we can see that the VGG-f net achieves the best results and that AlexNet achieves the worst results. The best result (90.5%) is achieved with the fully fine-tuned VGG-f net using SoftMax classification.

When we compare the CNN results with those of the comparison methods, we see that the fully fine-tuned VGG nets achieve clearly higher results than the comparison methods.

In Figure 4 we see two plots comparing the results on the validation data with the results on the training data during training of the nets (each 500th iteration). The nets where all trained on the 5th fold of the 5-fold cross validation. The upper plot shows the results of the fully fine-tuned nets during training and the bottom plot shows the results of the partly fine-tuned nets. In the upper plot we can see that the classification results of the four nets on the training data are all 100% or very close to 100%, whereat the results on the validation data are all close to 90 %. This shows that the fully fine-tuned nets are overfitted to the training data. We can observe that the differences between the results on the training and validation data are in general lower in case of the partly fine-tuned nets. This indicates that training only the fully connected layers leads to less overfitting than training all the layers of the nets. We can also observe that partly fine-tuning the nets requires more training iterations and leads to distinctly lower accuracies on the training data as for the fully fine-tuned nets. However, even though fully fine-tuning the nets leads to overfitting, the resulting nets are still better adapted to the the classification of celiac disease than the partly fine-tuned nets.

VI. CONCLUSION

In this work we showed that CNN transfer learning is very suited for the classification of celiac disease based on endoscopic image data. We used four CNNs that were pre-trained on the ImageNet database and used three different transfer learning strategies to classify the image patches of the

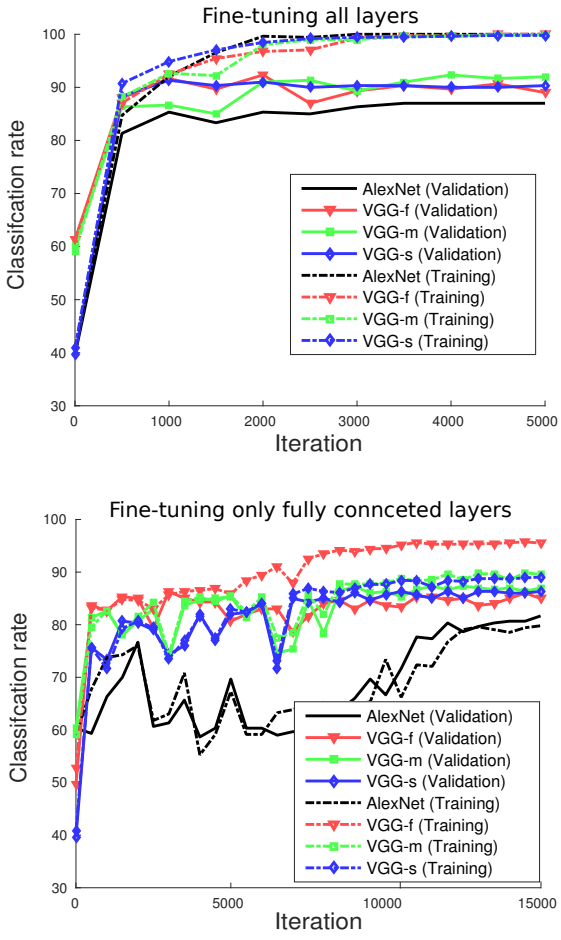


Figure 4. Training and Validation accuracies of the nets while training

celiac disease database. It turned out that fully fine-tuning the CNNs clearly achieves the highest classification accuracies, although the small amount of available training data leads to overfitting. The VGG-f net turned out to be the most suited network for the classification of celiac disease. The fully fine-tuned VGG nets outperformed the four state-of-the-art image representations.

References

[1] J. Jiang, P. Trundle, and J. Ren, "Medical image analysis with artificial neural networks," *Computerized Medical Imaging and Graphics*, vol. 34, no. 8, pp. 617 – 631, 2010.

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

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

[4] R. Zhu, R. Zhang, and D. Xue, "Lesion detection of endoscopy images based on convolutional neural network features," in *8th International Congress on Image and Signal Processing (CISP)*, Oct 2015, pp. 372–376.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[6] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *British Machine Vision Conference, BMVC 2014*, 2014.

[7] J. Yosinski, J. Clune, and H. Bengio, Yoshua andLipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, 2014, pp. 3320–3328.

[8] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: An astounding baseline for recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, ser. CVPRW '14, 2014, pp. 512–519.

[9] F. Biagi and G. Corazza, "Mortality in celiac disease," *Nature Reviews Gastroenterology and Hepatology*, vol. 7, pp. 158–162, 2010.

[10] A. Fasano, I. Berti, T. Gerarduzzi, T. Not, R. B. Colletti, S. Drago, Y. Elitsur, P. H. R. Green, S. Guandalini, I. D. Hill, M. Pietzak, A. Ventura, M. Thorpe, D. Kryszak, F. Fornaroli, S. S. Wasserman, J. A. Murray, and K. Horvath, "Prevalence of celiac disease in at-risk and not-at-risk groups in the united states: a large multicenter study," *Archives of internal medicine*, vol. 163, pp. 286–92, February 2003.

[11] F. Biagi, E. Rondonotti, J. Campanella, F. Villa, P. I. Bianchi, C. Klersy, R. D. Franchis, and G. R. Corazza, "Video capsule endoscopy and histology for small-bowel mucosa evaluation: A comparison performed by blinded observers," *Clinical Gastroenterology and Hepatology*, vol. 4, no. 8, pp. 998–1003, 2006.

[12] S. Niveloni, A. Fiorini, R. Dezi, S. Pedreira, E. Smecuol, H. Vazquez, A. Cabanne, L. A. Boerr, J. Valero, Z. Kogan, E. Maurino, and J. C. Bai, "Usefulness of videoduodenoscopy and vital dye staining as indicators of mucosal atrophy of celiac disease: assessment of interobserver agreement," *Gastrointestinal Endoscopy*, vol. 47, no. 3, pp. 223–229, 1998.

[13] N. Chand and A. A. Mihas, "Celiac disease: Current concepts in diagnosis and treatment," *Journal of Clinical Gastroenterology*, vol. 40, no. 1, pp. 3–14, January 2006.

[14] A. Gasbarrini, V. Ojetti, L. Cuoco, G. Cammarota, A. Migneco, A. Armuzzi, P. Pola, and G. Gasbarrini, "Lack of endoscopic visualization of intestinal villi with the immersion technique in overt atrophic celiac disease," *Gastrointestinal endoscopy*, vol. 57, pp. 348–351, 2003.

[15] F. Emura, Y. Saito, and I. H., "Narrow-band imaging optical chromocolonoscopy: advantages and limitations," *World J Gastroenterol*, vol. 14, no. 31, pp. 4867–4872, 2008.

[16] F. Valitutti, S. Oliva, D. Iorfida, M. Aloi, S. Gatti, C. M. Trovato, M. Montuori, A. Tiberti, S. Cucchiara, and G. Di Nardo, "Narrow band imaging combined with water immersion technique in the diagnosis of celiac disease," *Dig and Liver Dis*, vol. 46, no. 12, pp. 1099–1102, 2014.

[17] M. Gadermayr, S. Hegenbart, R. Kwitt, and A. Uhl, "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)*, Apr. 2016, pp. 355–359.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, 2015.

[19] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, "Learning multi-scale block local binary patterns for face recognition," in *Proceedings of the 2007 International Conference on Advances in Biometrics*, ser. ICB'07. Springer-Verlag, 2007, pp. 828–837.

[20] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in *Proceedings of the IEEE Digital Signal Processing Workshop, DSP '98*, Bryce Canyon, USA, Aug. 1998, pp. 9–12.

[21] F. Perronnin, Y. Liu, J. Sanchez, and H. Poirier, "Large-scale image retrieval with compressed fisher vectors," in *Proceedings of CVPR'10*, 2010, pp. 3384–3391.

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

[23] A. Vedaldi and K. Lenc, "Matconvnet – convolutional neural networks for matlab," in *Proceeding of the ACM Int. Conf. on Multimedia*, 2015, pp. 689–692.

[24] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.