Data Enrichment in Fine-Grained Classification of Aquatic Macroinvertebrates

Jenni Raitoharju*, Ekaterina Riabchenko*, Kristian Meissner†, Iftikhar Ahmad*,
Alexandros Iosifidis*, Moncef Gabbouj*, and Serkan Kiranyaz§
*Department of Signal Processing, Tampere University of Technology, Tampere, Finland
†Fresh Water Centre, Finnish Environment Institute, Jyväskylä, Finland
§Department of Electrical Engineering, Qatar University, Doha, Qatar
Email: *firstname.lastname@tut.fi, †firstname.lastname@ymparisto.fi, §mkiranyaz@qu.edu.qa

***Abstract*—The types and numbers of benthic macroinvertebrates found in a water body reflect water quality. Therefore, macroinvertebrates are routinely monitored as a part of freshwater ecological quality assessment. The collected macroinvertebrate samples are identified by human experts, which is costly and time-consuming. Thus, developing automated identification methods that could partially replace the human effort is important. In our group, we have been working toward this goal and, in this paper, we improve our earlier results on automated macroinvertebrate classification obtained using deep Convolutional Neural Networks (CNNs). We apply simple data enrichment prior to CNN training. By rotations and mirroring, we create new images so as to increase the total size of the image database sixfold. We evaluate the effect of data enrichment on Caffe and MatConvNet CNN implementations. The networks are trained either fully on the macroinvertebrate data or first pretrained using ImageNet pictures and then fine-tuned using the macroinvertebrate data. The results show 3-6% improvement, when the enriched data is used. This is an encouraging result, because it significantly narrows the gap between automated techniques and human experts, while it leaves room for future improvements as even the size of the enriched data, about 60000 images, is small compared to data sizes typically required for efficient training of deep CNNs.**

I. Introduction

Preserving good water quality and sustainable water resource management are important for human well-being both locally and globally. To better understand the factors threatening water quality and to use protective measures as efficiently as possible, it is important to monitor aquatic communities, because physicochemical water quality assessment cannot alone provide a sufficient understanding of the impacts of various human activities on aquatic ecosystems. Therefore, targeted environmental legislation, such as the EU Water Framework Directive 2000/60/EC (WFD), requires monitoring also several biological quality elements in freshwater ecological status assessment. Among these elements, benthic macroinvertebrates are one of the most commonly monitored groups [1].

Benthic macroinvertebrates are bottom dwelling animals sensitive to physical and chemical changes in their habitat [2]. They are relatively long-lived, semi-stationary in their larval stages, and easy to collect. Shifts in the number of different macroinvertebrates can reveal changes in the ecological status of a water body. Macroinvertebrates can be used to evaluate consequences of past actions and to direct future management actions. Thus, it is critical that the collected macroinvertebrate specimens are identified correctly. The identification is not easy as there are hundreds of different macroinvertebrate taxa in the boreal fresh waters of Europe and specimens from related taxa can be similar, while specimens from the same taxa can exhibit significant differences. This is illustrated in Fig. 1, where four specimens from two different taxa are shown. In classification terms, there are high intra-class and low inter-class variations and fine-grained classification is required.

Figure 1: Four samples from *Ephemerella aroni (aurivillii)* (top) and *Leuctra sp* (bottom) classes.

Human expert taxonomic microscopic identification is costly and time intensive. Furthermore, human identifications have been repeatedly shown to be less reliable than assumed (e.g. [3], [4]). Therefore, the need for developing automatic taxa identification methods that could eventually replace a large part of the required human effort is imminent. In our recent study [5], we showed that Convolutional Neural Networks (CNNs) can achieve near human accuracy in identification of Finnish river macroinvertebrate benthos and beat the traditional classification approaches using hand-crafted features. In this study, we further improve the results by applying data enrichment prior to CNN training.

II. Related Work

Macroinvertebrate identification requires techniques capable of fine-grained classification. In recent years, CNNs have been shown to perform well on several fine-grained classification tasks such as car model classification [6], [7], [8], bird species recognition [9], [8], plant identification [10], or plankton classification [11]. However, in the aforementioned tasks, different classes are generally easily distinguished —even if not recognized —also by non-experts. For macroinvertebrates, this is not the case as related taxa are sub-classes of the same family and the morphological differences between the taxa are often too subtle for a non-expert to detect.

Earlier efforts to automatically classify benthic macroinvertebrates have used different classifiers exploiting hand-crafted features [12], [13], [14], [15]. In [12], [13], [15], the number of taxa was limited to 8-9 and promising identification results with over 90% accuracy were obtained. In [14], a larger dataset including 35 taxa and 6814 images was considered and the best automatic classification accuracy was slightly above 80%. In our recent work [5], we applied the current state-of-the-art classifier, deep CNN, for macroinvertebrate classification. That study showed that features learned by a deep CNN can outperform previously used hand-crafted features. Furthermore, we obtained the best classification results using a CNN which was pretrained using completely irrelevant ImageNet [16] images and only fine-tuned using the macroinvertebrate data. The dataset considered in [5] contained 29 taxa and 11832 images and the best average classification accuracy obtained was about 85%. The same database is used in this study.

III. Data Description

A. Database Acquisition

The benthic macroinvertebrates evaluated in this study were originally collected from Finnish river bottoms. Once carefully cleaned from debris, a human expert identified the samples using a microscope. Depending of the difficulty of the taxa identification, the macroinvertebrates were classified with a varying taxonomic resolution. Most taxa were identified to the species level, but, if this was not reliably possible, a coarser level was applied (in Table I, genus level is denoted with ”sp.” and non-italic font denotes an even coarser level). Overall, we considered 29 different macroinvertebrate taxa in this study.

After identification, the macroinvertebrates were imaged using a simple setup shown in Fig. 2. The macroinvertebrates were dropped one-by-one to a test-tube filled with alcohol. During their fall to the bottom of the test tube, the specimens were pictured using two separate cameras viewing the test tube from perpendicular angles as shown in the figure. Thus, for each macroinvertebrate specimen, two simultaneous images from different viewing perspectives were taken. The images were in PNG format and their size (at this point prior to preprocessing) varied from 640x480 to 1280x960 pixels. We intend to later publish the data used in this study as a part of a larger publicly available benchmark database of benthic macroinvertebrates.

B. Preprocessing

To improve the quality of the images used for CNN training, all the images underwent the same preprocessing steps. Some images were completely discarded due to their poor quality,

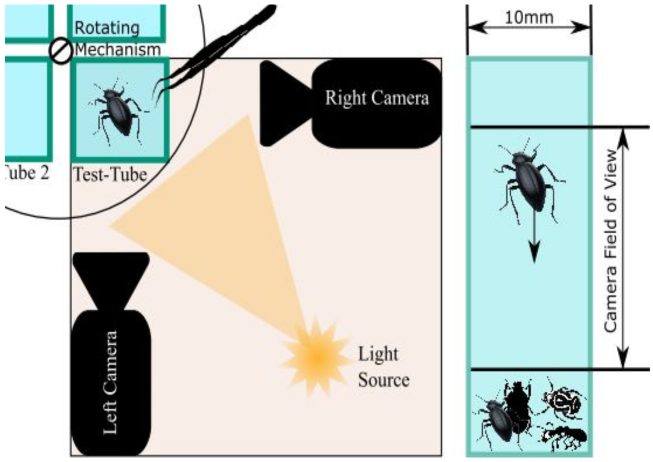


Figure 2: Imaging setup used in data acquisition

while the remaining images were cropped to focus on the macroinvertebrates and to remove artifacts mainly appearing close to image borders. In order to crop without losing important information, we applied Otsu’s method [17] and analysis of connected components to produce binary masks. A binary mask was used to define the outer edges of each specimen and an additional margin of at least 20 pixels was added before cropping out a square region. Finally, the cropped patch was scaled to the size of 256x256 pixels. The cropping is illustrated in Fig. 3, where the original images and the cropped versions along with the corresponding binary masks are shown for two macroinvertebrate specimens.

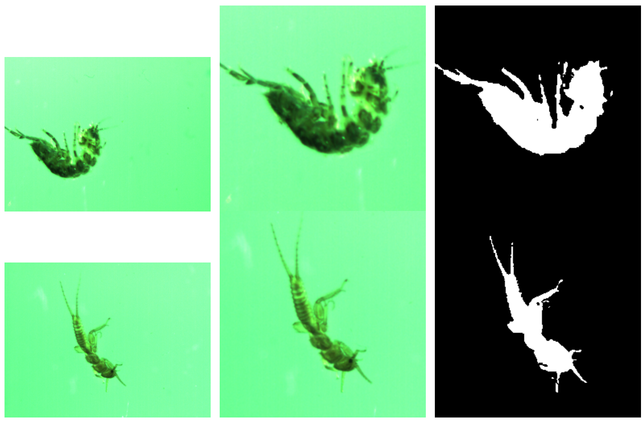


Figure 3: From left to right: original image, cropped image, corresponding binary mask. Top row - *Asellus aquaticus*, bottom row - *Nemoura*.

The final numbers of images for each macroinvertebrate taxa are given in Table I. The smallest class, *Atherix ibis* is represented by only 230 images, while the largest class, *Ephemerella aroni (aurivillii)*, consists of 577 images. The total number of original images is 11832. We divided the data randomly into training, testing and validation sets as follows: training - 50%, testing - 30%, and validation - 20%.

To evaluate the effect of data partition, we created 10 different random data partitions. Both images of the same instance taken simultaneously with two different cameras as explained above were always placed into the same set (training, testing, or validation). In this study, the validation set was not exploited, but we adopt the data partitions from [5] in order to enable direct comparison of the results.

C. Data Enrichment

CNNs have millions of parameters and, accordingly, millions of images are required to learn the parameters without considerable overfitting. A common way to combat overfitting is to artificially enrich the data. In the standard *AlexNet* implementation [18], the size of the training data has been increased 2048-fold by translating and mirroring the original images. Even more new images are created by altering the intensities of RGB channels. This data enrichment effectively helps avoiding overfitting and, thus, allows training a substantially larger network. The data enrichment has been successfully applied also for fine-grained classification, e.g. for facial expression recognition in [19].

The available macroinvertebrate data consist of only 11832 images (half of them used for training), which is significantly too few for training deep CNN. Therefore, it is not surprising that in [5] the best results are obtained using a network pretrained on ImageNet images. While these images are completely irrelevant to the current application, the pretraining can help to learn features and avoid overfitting caused by the small size of the macroinvertebrate dataset. In this study, we apply only a limited number of simple enrichment techniques. From each original image, we create 3-8 new images depending on the original number of images for the corresponding taxon. The expansion factors for each taxon are specified in Table I. We aim at having about 2000 images for each class and total number of images is increased to 62960. Also the number of enriched images remains small for efficient CNN training, but we show that already this enrichment can significantly narrow the gap between the automated CNN classifier and human experts, which usually have a 90%-95% identification accuracy.

In this study, we apply only rigid transformations: rotation and mirroring. We start by horizontal and vertical mirroring and continue by random rotations until the target expansion factor is reached. An example of the data enrichment for expansion factor 5 is shown in Fig. 4. It remains a topic for the future work to apply different transformations such as addition of noise, jittering, nonrigid transformations, or cropping of some body parts (e.g. head or tail).

IV. Experimental results

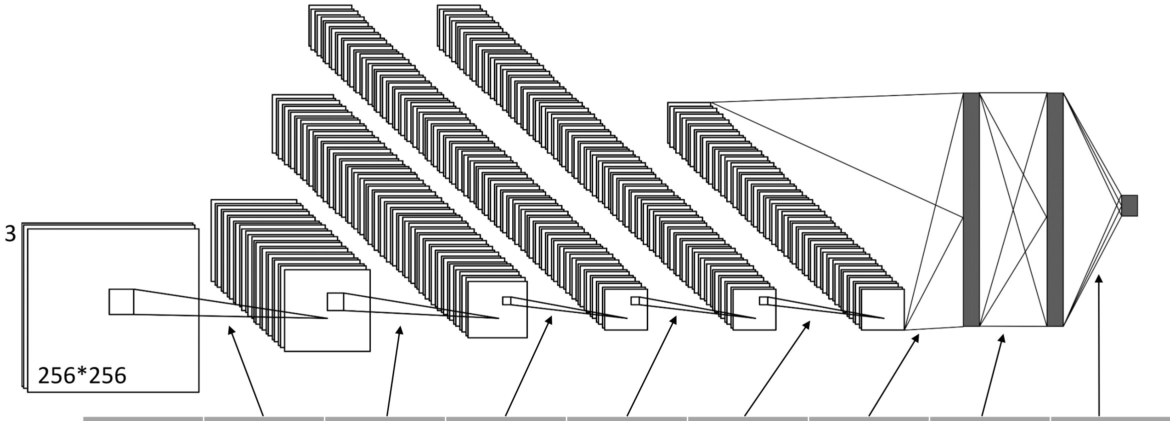
A. Applied CNNs

We experimented on two different publicly available implementations of deep CNN, namely MatConvNet [20] and Caffe [21]. MatConvNet is a Matlab toolbox, while Caffe is a C++ library with Python and Matlab bindings. For both networks, we used an architecture close to the landmark *AlexNet* model

Figure 4: Example of data enrichment for expansion factor 5. Top left - original image, top center and right - horizontally and vertically mirrored images, bottom - images rotated by an arbitrary angle

Table I: Considered macroinvertebrate taxa, original number of images in each taxon, and the corresponding expansion factor

#	Macroinvertebrate taxon	Orig. no. of images	Expansion factor
1	*Ameletus inopinatus*	343	5
2	*Asellus aquaticus*	447	4
3	*Atherix ibis*	230	8
5	*Baetis niger*	455	4
4	*Baetis rhodani*	468	4
6	Ceratopogonidae	322	5
7	*Dicranota* sp.	367	5
8	*Elmis aenea*	468	4
9	*Elmis aenea* adult	378	4
10	*Ephemerella aroni (aurivillii)*	577	3
11	*Habrophlebia* sp.	458	4
12	*Hemerodromia* sp.	280	7
13	*Heptagenia dalecarlica*	409	4
14	*Hydraena* adult	436	4
15	*Isoperla* sp.	460	4
16	*Itytrichia lamellaris*	428	4
17	*Leptophlebia* sp.	480	4
18	*Leuctra* sp.	378	5
19	*Limnius volckmari* adult	395	4
20	*Micrasema gelidum*	417	4
21	*Micrasema setiferum*	372	5
22	*Nemoura* sp.	414	4
23	*Oulimnius tuberculatus*	465	4
24	*Oxyethira* sp.	438	4
25	*Philopotamus montanus*	330	5
26	Psychodiidae	408	4
27	*Protonemura* sp.	387	5
28	Simuliidae	418	4
29	*Taeniopteryx nebulosa*	404	4



	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7	Layer 8
Type	conv+max +norm	conv+max +norm	conv	conv	conv+max	fully connected	fully connected	fully connected
Channels	96	256	384	384	256	4096	4096	29
Filter size	11x11	5x5	3x3	3x3	3x3			
Input size	227*227	55*55	27*27	13*13	13*13			

Figure 5: AlexNet model as implemented in Caffe

[18] illustrated in Fig. 5. There are eight layers in total. The first seven layers, five convolutional and two fully-connected layers, can be considered as a high-level feature extractor. The output layer then makes the classification decisions based on the extracted features. For convolutional layers, the number of filters (channels) varies along with the filter size. Only 2-dimensional filter sizes are shown in the figure, but the filters in fact extend through all input channels. For example, in the second layer having 96 input channels, the number of parameters required for a single filter is 5x5x96. For the first layer, the number of input channels is three, because color images are considered. The convolution does not change the image size, but it may be followed by max-pooling to reduce the image size. Note that the size of the input images (256x256) is different from the input size (227x227) given in the figure. Generally, the AlexNet model crops 10 patches of the original figure and the classification is done based on their average. There are minor differences in the implementation details between Caffe and MatConvNet. For Caffe, max-pooling is carried out before normalization, while the opposite is true for MatConvNet and original AlexNet model. Also the cropping procedures to obtain 227x227 patches are different. Further implementation details can be found in the original works [18], [20], [21].

We first trained both networks using only the macroinvertebrate data. For Caffe, the maximum number of training iterations was set to 10000/15000 and a minibatch size of 64 was used. The number of training epochs can be computed by dividing the number of iterations by the minibatch size. For MatConvNet, 60 training epochs were run. With Caffe, we also experimented with a network pretrained using ImageNet [16] images. This network was then fine-tuned using the macroinvertebrate data (5000 iterations).

B. Evaluation

The trained networks were evaluated based on their ability to correctly classify the macroinvertebrate specimens into one of the taxa given in Table I. As always two concurrent images from different viewing perspectives exist for each instance, the final identification decision was based on the both images. For the enriched dataset, classification was based on all the images generated from the original pair. If the labels obtained from network outputs for different images of the same instance agreed, this label was naturally assigned to the instance. However, when the labels disagreed, the final label was based on voting, where also the confidence of the outputs was considered. Finally, the classification accuracy was computed as the proportion of correctly classified images in the testing set.

C. Comparison Results

Table II presents comparative results between networks trained using the original and enriched data. The results are given separately for each database partition. The results using the original training set on Caffe, indicated by asterisks, have been originally published in [5].

The results clearly indicate that enriching the training set before training CNNs for classification of macroinvertebrates is beneficial. The classification accuracy has improved for every data partition on both networks, with or without pre-training. The average improvement varies from 2.9% obtained of pretrained Caffe architecture to 6,28 % on MatConvNet. Also the deviation of results has clearly decreased.

Table II: Classification accuracies obtained by training with original and enriched training sets

	MatConvNet		Caffe		Caffe pretrained	
#	orig.	enriched	orig.*	enriched	orig.*	enriched
1	81.28	84.27	77.41	79.44	85.36	88.27
2	77.58	86.95	77.47	80.98	86.35	89.04
3	80.15	86.07	75.93	80.87	86.13	87.83
4	77.19	85.53	74.56	80.43	84.76	88.65
5	81.20	85.47	79.06	82.02	86.13	88.65
6	78.56	85.53	77.69	80.48	85.42	88.27
7	79.22	85.53	76.97	78.84	86.02	88.32
8	80.92	84.81	76.70	79.99	85.47	89.09
9	76.48	84.76	76.26	80.26	84.38	87.99
10	78.84	85.31	77.69	81.47	86.40	89.31
avg	79.14	85.42	76.97	80.48	85.64	88.54
std	1.72	0.74	1.22	0.93	0.69	0.49

V. Conclusion

To improve the accuracy of CNNs in automated classification of benthic river macroinvertebrates, we applied rigid transformations, rotation and mirroring, to the original images of macroinvertebrates to generate more images thus enriching the data. The effect of data enrichment was evaluated on two publicly available CNN implementations, Caffe and MatConvNet. The network training was either fully performed on the macroinvertebrate data or first pretrained using irrelevant ImageNet pictures and then only fine-tuned using the macroinvertebrate data. The results confirmed the usefulness of data enrichment observed on other application domains: using enriched training data, the accuracy of fine-grained classification of benthic macroinvertebrates improved by 3-6%.

The results are encouraging. Data enrichment proved to be an easy way to significantly narrow the performance gap between automated methods and human experts. This is despite the fact that even the size of the enriched database, 62960 images, remains small compared to the data sizes typically required to efficiently train deep CNNs. Thus, the results suggest that, in the near future, the performance of CNNs can indeed catch up human expert performance through acquisition of larger labeled macroinvertebrate datasets and by application of more advanced enrichment techniques combined with substantially larger data expansion factors.

Acknowledgment

The authors would like to thank the Academy of Finland for the grants no. 288584 and 289104 funding the DETECT consortium's project (Advanced Computational and Statistical Techniques for Biomonitoring and Aquatic Ecosystem Service Management).

References

[1] S. Birk, W. Bonne, A. Borja, S. Brucet, A. Courrat, S. Poikane, A. Solimini, W. Van de Bund, N. Zampoukas, and D. Hering, "Three hundred ways to assess europe's surface waters: An almost complete overview of biological methods to implement the water framework directive," *Ecological Indicators*, vol. 18, pp. 31–41, 2012.

[2] Y. Lin, Q. Chen, K. Chen, and Q. Yang, "Modelling the presence and identifying the determinant factors of dominant macroinvertebrate taxa in a karst river," *Environmental Monitoring and Assessment*, vol. 188, no. 6, pp. 1–11, 2016.

[3] P. Haase, S. U. Pauls, K. Schindehütte, and A. Sundermann, "First audit of macroinvertebrate samples from an eu water framework directive monitoring program: human error greatly lowers precision of assessment results," *Journal of the North American Benthological Society*, vol. 29, no. 4, pp. 1279–1291, 2010.

[4] P. F. Culverhouse, "Human and machine factors in algae monitoring performance," *Ecological Informatics*, vol. 2, no. 4, pp. 361–366, 2007.

[5] E. Riabchenko, K. Meissner, I. Ahmad, A. Iosifidis, V. Tirronen, M. Gabbouj, and S. Kiranyaz, "Learned vs. engineered features for fine-grained classification of aquatic macroinvertebrates," in *Accepted to International Conference on Pattern Recognition (ICPR)*, 2016.

[6] J. Krause, T. Gebru, J. Deng, L.-J. Li, and L. Fei-Fei, "Learning features and parts for fine-grained recognition," in *22dn International Conference on Pattern Recognition (ICPR)*, pp. 26–33, Aug 2014.

[7] L. Yang, P. Luo, C. C. Loy, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3973–3981, June 2015.

[8] J. Krause, H. Jin, J. Yang, and L. Fei-Fei, "Fine-grained recognition without part annotations," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5546–5555, June 2015.

[9] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, *Part-based R-CNNs for fine-grained category detection*, pp. 834–849. 2014.

[10] P. B. H. Goau, A. Joly, "Plant identification in an open world (lifeclef2016)," tech. rep., 2016.

[11] H. Lee, M. Park, and J. Kim, "Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning," in *IEEE International Conference on Image Processing (ICIP)*, pp. 3713–3717, Sept 2016.

[12] D. A. Lytle, G. Martínez-Muñoz, W. Zhang, N. Larios, L. Shapiro, R. Paasch, A. Moldenke, E. N. Mortensen, S. Todorovic, and T. G. Dietterich, "Automated processing and identification of benthic invertebrate samples," *Journal of the North American Benthological Society*, vol. 29, no. 3, pp. 867–874, 2010.

[13] S. Kiranyaz, T. Ince, J. Pulkkinen, M. Gabbouj, J. Ärje, S. Kärkkäinen, V. Tirronen, M. Juhola, T. Turpeinen, and K. Meissner, "Classification and retrieval on macroinvertabrate image databases," *Computers in Biology and Medicine*, vol. 41, pp. 463–472, Jul. 2011.

[14] J. Ärje, S. Kärkkäinen, T. Turpeinen, and K. Meissner, "Breaking the curse of dimensionality in quadratic discriminant analysis models with a novel variant of a bayes classifier enhances automated taxa identification of freshwater macroinvertebrates," *Environmetrics*, vol. 24, no. 4, pp. 248–259, 2013.

[15] H. Joutsijoki, K. Meissner, M. Gabbouj, S. Kiranyaz, J. Raitoharju, J. Ärje, S. Kärkkäinen, V. Tirronen, T. Turpeinen, and M. Juhola, "Evaluating the performance of artificial neural networks for the classification of freshwater benthic macroinvertebrates," *Ecological Informatics*, vol. 20, pp. 1–12, Mar. 2014.

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[17] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[19] A. Maronidis, D. Bolis, A. Tefas, and I. Pitas, "Improving subspace learning for facial expression recognition using person dependent and geometrically enriched training sets," *Neural Networks*, vol. 24, no. 8, pp. 814 – 823, 2011. Artificial Neural Networks: Selected Papers from {ICANN} 2010.

[20] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *ACM International Conference on Multimedia*, 2015.

[21] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for

fast feature embedding," in *International Conference on Multimedia*, pp. 675–678, 2014.