ASSESSMENT OF LABORATORY MOUSE ACTIVITY IN VIDEO RECORDINGS USING DEEP LEARNING METHODS

Marcin Kopaczka, Daniel Tillmann, Lisa Ernst, Justus Schock, René Tolba and Dorit Merhof

Institute of Imaging and Computer Vision, RWTH Aachen University, Germany

ABSTRACT

Analysis of laboratory animal behavior allows assessment of animal wellbeing. We present a method for the classification of different activities of laboratory mice by analyzing video clips using three deep learning methods. Animals placed in observation cages are filmed and short video clips are labelled as belonging to one of five defined behaviors. Subsequently, three different methods based on convolutional neural networks (CNNS) are applied to classify the clips. The best performing method - a two-stream network that analyzes individual frames as well as the video's optical flow - achieves an accuracy of 86.4%, including detection of important behavioral patterns such as self-grooming. These results show that the presented analysis protocol allows automated assessment of animal behavior by algorithmic analysis of videos of mice on observation boxes.

***Index Terms*—** Mouse grimace scale, Severity Assessment, convolutional neural networks

1. INTRODUCTION AND PREVIOUS WORK

Experiments with laboratory animals are strictly regulated. The 3R principle (replace, reduce, refine), first published in [1] and implemented by current standards and guidelines for laboratory animal experiments, sets strict requirements on the planning and execution of lab animal trials. Following the 3R principle requires that experiments on living animals must be fully *replaced* by other experiments that do not require animal handling wherever possible. Should experiments on animals still be necessary for the given research task, then the number of animals undergoing trials must be *reduced* to the minimum number of animals required for meaningful results. For these remaining animals, the experiments must be *refined* to minimize animal disstress and suffering. To ensure that these guidelines are implemented in practice, a number of scoring methods for assessing animal disstress during experiments has been established. These methods can be roughly subdivided into two categories: device-based and observation-based. Device-based methods require special experimental equipment to assess the physical and neurological status of animals. Examples include the rotarod [3], where rodents are placed on a rotating cylinder and the time that the animal successes to remain on the rotating device is measured, or the balance beam [4], where the animals need to cross the gap between two platforms on a narrow beam and the time required to complete this task is measured. Due to their quantitative output, device-based methods have the advantage of minimizing inter-observer variability. In observation-based methods, scores focusing on animal behavior are assessed by the laboratory staff. Examples include the analysis of nesting and burrowing behavior [5], animal tracking in open field experiments [6] or the mouse grimace scale (MGS), in which the facial expressions of animals are scored as an indicator of disstress [7]. While they require no or minimal additional equipment the downslide of these methods is that they require time-consuming manual assessment. To this end, a number of methods for automated video-based analysis has been developed. Notable examples include the Mice Profiler [8], the Noldus software system and early systems for automated MGS scoring [9]. They all aim at automating the analysis and producing quantitative results to reduce inter-reader variability and human workload.

Fig. 1. A still frame showing a grooming animal (left) and a calm animal (right).

In this work, we focus on activity assessment in an observation box originally developed for mouse grimace scale analysis. Instead of analyzing the facial expressions of mice according to MGS, we classify different animal behaviors that may serve as indicators for animal disstress. We implemented a set of different deep neural networks to distinguish calm from active animals and to additionally classify specific activities such as self-grooming, as anomalies in the self-grooming behavior of mice are both an indicator of disstress and are also correlated with different neuroscientific pathologies.

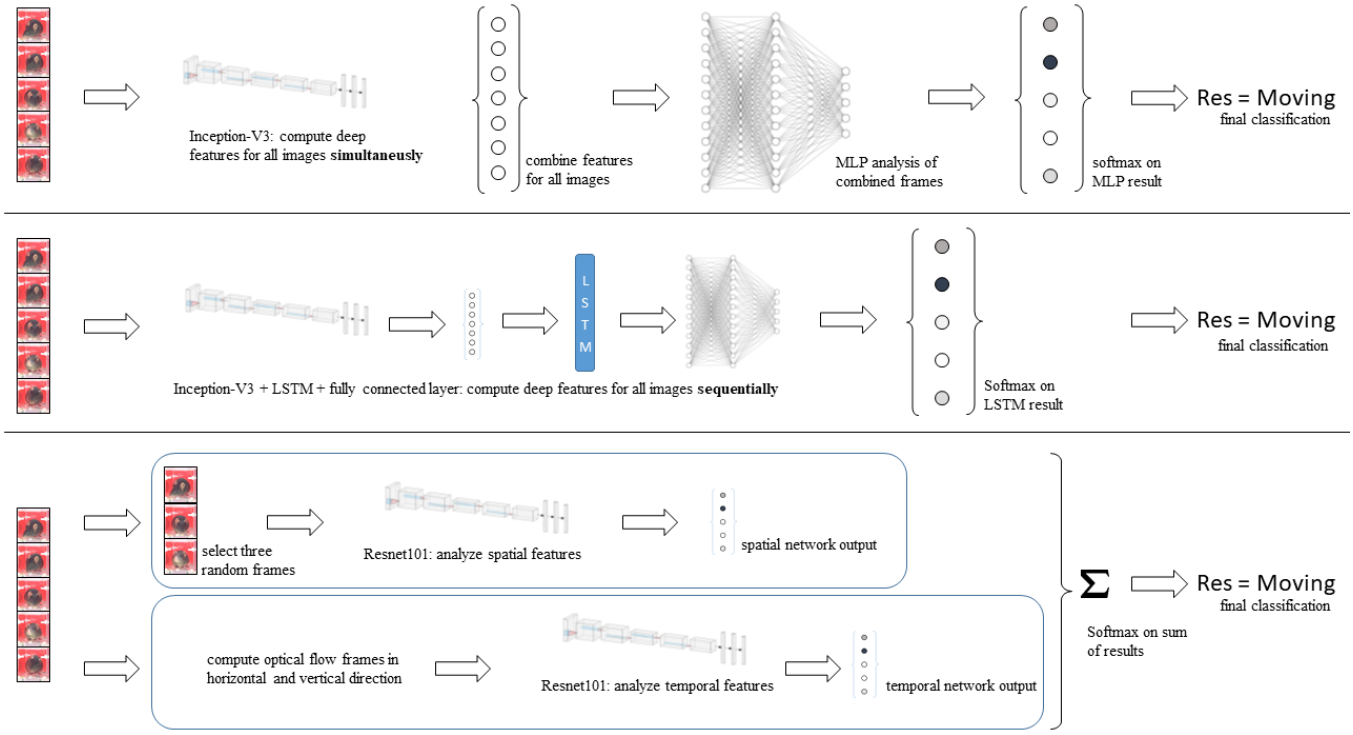

Fig. 2. The three approaches implemented in our work. From top to bottom: CNN+MLP, CNN+LSTM, Two-Stream Network.

2. MATERIALS AND METHODS

Our samples were extracted from videos recorded for mouse grimace scale assessment of laboratory mice undergoing experiments at an animal laboratory. The setup was designed by the laboratory experts for efficient MGS assessment of up to 4 mice in individual boxes.

We cropped video clips of single animals with a duration between 3 and 8 seconds showing one of the following behaviors:

- Resting - the animal is calm, showing no or minimal movement. The animal is facing the camera.
- Grooming - self-grooming while facing the camera.
- Turning around - turning by 180 degrees towards the camera or away from it at least once during the clip.
- General movement - actions not belonging to one of the above categories, for example sniffing around or cage exploration without the animal turning it's back towards the camera.
- Turned away - The animal's back is facing the camera for the entire clip duration, regardless of animal movement.

Each video clip was showing exactly one of the described behaviors with no mixed behaviors in the dataset such as grooming and turning away afterwards.

Subsequently, we implemented three methods for action recognition, with each method showing a different approach using deep learning techniques. Method 1 combines a convolutional neural network (CNN) with a multilayer perceptron (MLP). The convolutional network - we used the widely used Inception-V3 network [10] - receives a set of n input frames sampled equidistantly from the input video clip and processes each frame independently, returning a 2048-entry feature vector. Subsequently, the feature vectors of all 30 frames are combined into a (n x 2048)-entry vector that is passed to the MLP. The MLP itself consists of two fully connected layers with 512 neurons each, both followed by a dropout layer. The final prediction is made using a softmax layer with one entry for each of the aforementioned classes.

Method 2 also makes use of the Inception-V3 CNN architecture for feature extraction, however the resulting feature vectors are not combined into a single vector but instead fed sequentially into a LSTM network. This method takes advantage of the fact that LSTMs contain an explicit temporal memory allowing them processing time-series input while preserving the temporal structure of the data. We used a single LSTM cell followed by a 512-neuron dense layer and a 0.5 dropout layer and the same softmax output as above.

Method 3 combines two CNNs in a two-stream network

(TSN) that allow analysis of video streams using regular CNNs [11]. In our case, we used the ResNet101 architecture [12] for both networks. The first network analyzes the spatial properties of the data by picking three random frames from the video clip and combining them into a single 9-channel input image (3 images with 3 color channels each). The second network analyzes the temporal structure of the data based on optical flow analysis. We computed the optical flow seperately in horizontal and vertical direction and extracted 30 equidistant optical flow frames from each video clip. These were combined into a single 60-channel image and processed by the temporal network. The output of each of the networks is a feature vector with class probabilities for each of the given behavior classes, the final prediction is made by applying softmax to the sum of both outputs.

Note that the actual length of the video clips does not need to be constant. All methods are able to process video clips of arbitrary length since they sample single frames from the input videos regardless of their length.

3. EXPERIMENTS AND RESULTS

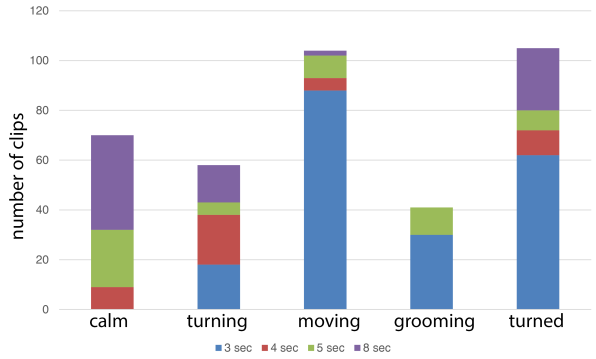

Fig. 3. Class distribution and sequence lengths of the clips in the dataset.

For evaluation, a total of 378 clips were extracted from a set of 12 videos: 70 clips of resting animals, 41 grooming clips, 58 clips for turning around, 104 clips for general movement and 105 clips with animals turning their back at the camera. The clips had a duration between 3 and 8 seconds(Fig. 3). The clips were split into five groups containing about 20% of the videos each and the networks were trained using 5-fold cross validation with four sets for training and the remaining set for testing. The clips were resized to 224 pixels width and height since this is the input image size that the used convolutional networks were designed for.

All networks were evaluated with a different number of extracted frames per clip. We evaulated the performance for 1, 5, 10, 20, 30, 40 and 50 extracted frames by analyzing the percentage of correctly recognized videos (Top1 accuracy) and show the results in Fig. 4. It can be seen that the best

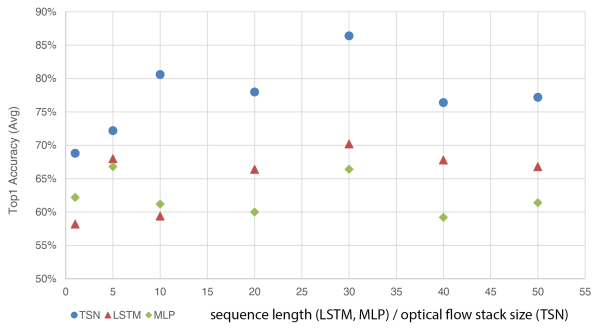

Fig. 4. Top1 accuracies of the three presented networks depending on the number of analyzed frames per clip.

overall performance is achieved at 30 frames where both the TSN and the LSTM achieve their best results and the MLP is extremly close to its absolute maximum which is at 5 frames per clip (66.4% vs. 66.8%). Therefore, all further presented results are given for networks analyzing 30 frames of the clip.

Fig. 5, 6 and 7 show the detailed confusion matrices for the three networks. The MLP method shows strong detection rates for the general movement class, however with a large number of false classifications for the self-grooming clips where self-grooming is systematically wrongly identified as general movement. The LSTM network shows similar behavior to the MLP with better overall performance, however the misclassification of grooming animals is still at 50%. Animals that are turned away are classified with a very high score of 0.97. The TSN shows the best overall performance with a perfect identification of calm animals and very good detection rates for distinguishing moving from grooming animals, a task that had been very challenging for the other methods. A run time comparison of all methods is shown in Fig. 8. All methods have run times between 200 and 420 ms for a multi-second clip, making real-time applications possible. Depending on the method, most time is either spent on feature computation in the CNNs (for LSTM and MLP) or for optical flow computation when using TSNs.

4. DISCUSSION AND CONCLUSION

The implemented methods allow activity recognition from video clips, with TSNs yielding the best performance with a Top1 accuracy of 86.4%. Out of the five predefined classes, self-grooming and general motion are the most hardest to distinguish from each other, while calm animals and those which are always turned away from the camera can be identified with excellent accuracy. Future work will include analysis of video clips of different mice strains or animals undergoing medical treatment for automated assessment of behavioral changes as well as using the classifiers for selecting clips of animals not in motion for automated MGS scoring.

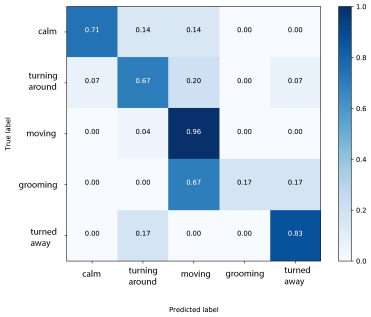

Fig. 5. MLP confusion matrix

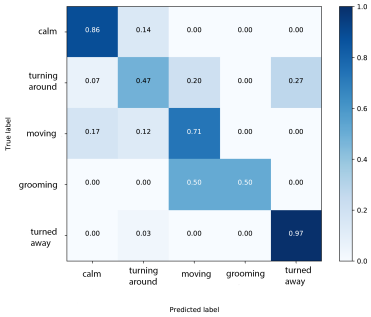

Fig. 6. LSTM confusion matrix

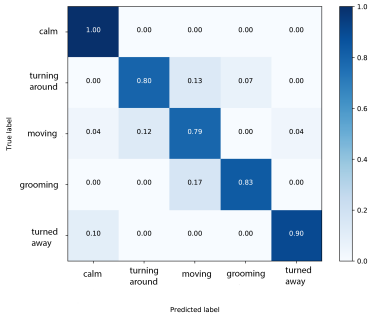

Fig. 7. TSN confusion matrix

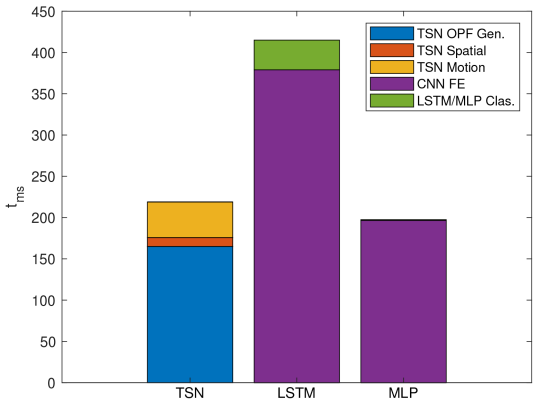

Fig. 8. Run times of the implemented methods for predicting one clip.

5. REFERENCES

[1] William Moy Stratton Russell and Rex Leonard Burch, "The principles of humane experimental technique," 1959.

[2] BJ Jones and DJ Roberts, "The quantitative measurement of motor inco-ordination in naive mice using an accelerating rotarod," *Journal of Pharmacy and Pharmacology*, vol. 20, no. 4, pp. 302–304, 1968.

[3] BJ Jones and DJ Roberts, "A rotarod suitable for quantitative measurements of motor incoordination in naive mice," *Naunyn-Schmiedeberg's Archives of Pharmacology*, vol. 259, no. 2, pp. 211–211, 1968.

[4] Tinh N Luong, Holly J Carlisle, Amber Southwell, and Paul H Patterson, "Assessment of motor balance and coordination in mice using the balance beam," *Journal of visualized experiments: JoVE*, , no. 49, 2011.

[5] Paulin Jirkof, "Burrowing and nest building behavior as indicators of well-being in mice," *Journal of neuroscience methods*, vol. 234, pp. 139–146, 2014.

[6] Michael L Seibenhener and Michael C Wooten, "Use of the open field maze to measure locomotor and anxiety-like behavior in mice," *Journal of visualized experiments: JoVE*, , no. 96, 2015.

[7] Dale J Langford, Andrea L Bailey, Mona Lisa Chanda, Sarah E Clarke, Tanya E Drummond, Stephanie Echols, Sarah Glick, Joelle Ingrao, Tammy Klassen-Ross, Michael L LaCroix-Fralish, et al., "Coding of facial expressions of pain in the laboratory mouse," *Nature methods*, vol. 7, no. 6, pp. 447, 2010.

[8] Fabrice De Chaumont, Renata Dos-Santos Coura, Pierre Serreau, Arnaud Cressant, Jonathan Chabout, Sylvie Granon, and Jean-Christophe Olivo-Marin, "Computerized video analysis of social interactions in mice," *Nature methods*, vol. 9, no. 4, pp. 410, 2012.

[9] Marcin Kopaczka, Lisa Ernst, Justus Schock, Arne Schneuing, Alexander Guth, René H. Tolba, and Dorit Merhof, "Introducing cnn-based mouse grim scale analysis for fully automated image-based assessment of distress in laboratory mice," in *Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM)*, 2018.

[10] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[11] Karen Simonyan and Andrew Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in neural information processing systems*, 2014, pp. 568–576.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.