Automated Mouse Behavior Recognition using LSTM and TCN Networks.

Xavier de Juan Pulido

Fundamental Principles of Data Science Master's Thesis

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

July 15, 2022

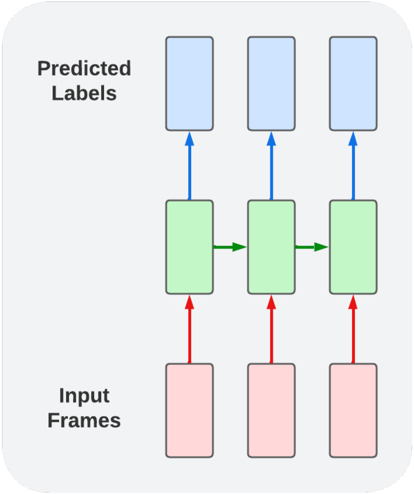
Índex

Introduction

Motivation

- Project initiated by Dr. Mercè Masana, Professor of the Department of Biomedicine from the Faculty of Medicine and Health Sciences in the University of Barcelona.
- Research studies in neurosciences employ mice for their experiments.
- These trials require a significant amount of behavior analysis.
- **Main goal**: automate the frame tagging by using computer vision and deep learning.

The Problem



- Label images by their content.
- Animal motion recognition task
- In Computer Vision: multiclass video classification on frame level
- Sequence processing many to many problem.

Related Work

Architectures already applied

- *Kopaczka et al., 2019*: Two-Stream Convolutional Networks.
- *Bohnslav et al., 2020*: 3D Convolutional Networks.
- *Ngoc Giang et al., 2019*: Two-Stream I3D Convolutional Networks.
- *Zhang, Yang, and Wu, 2019*: CNN+LSTM architecture.

Methodology

Data

- 17 mice recordings with the annotated behaviors: grooming, mid rearing and wall rearing.

(a) Grooming

(b) Mid Rearing

(c) Wall Rearing

Data Transformations

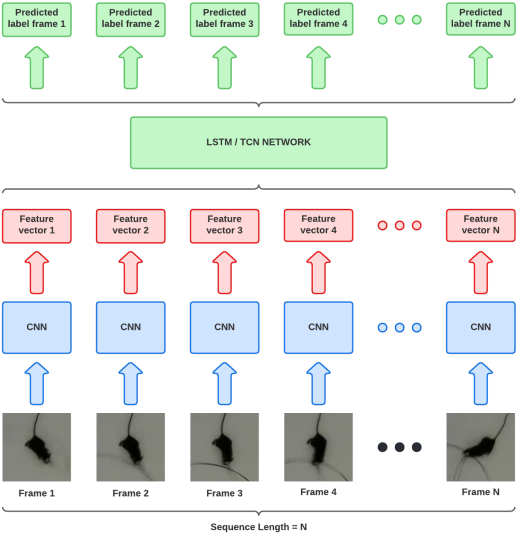
- Cropping using DeepLabCut™software.
- Reduce the video length in smaller sequences.

Methods

- ResNet + LSTM
- InceptionResNet + LSTM
- ResNet + TCN
- InceptionResNet + TCN

Training Strategy

- Neural Architecture and Hyperparameter search: cross-validation across each video.

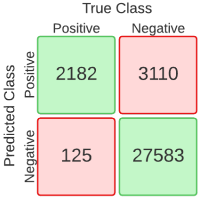


Evaluation Metrics

- Confusion Matrix: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN)
- Binary Accuracy: $\frac{TP+TN}{TP+TN+FP+FN}$
- Precision: $\frac{TP}{TP+FP}$. Proportion of positive identifications that were actually correct.
- Recall: $\frac{TP}{TP+FN}$. Proportion of actual positives that were identified correctly.
- Area under the precision-recall curve.

Results: Grooming

	ResNet+LSTM	Inc.ResNet+LSTM	ResNet+TCN	Inc.ResNet+TCN
Binary Accuracy	0.902	0.909	**0.950**	0.942
Precision	0.412	0.426	**0.681**	0.577
Recall	**0.946**	0.876	0.537	0.635
AUC PRC	0.679	**0.690**	0.588	0.586



Experiment 1 Confusion Matrix

Experiment 2 Confusion Matrix

Experiment 3 Confusion Matrix

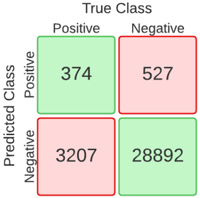
Experiment 4 Confusion Matrix

Results: Mid Rearing

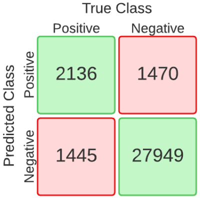
	ResNet+LSTM	Inc.ResNet+LSTM	ResNet+TCN	Inc.ResNet+TCN
Binary Accuracy	0.897	0.902	0.887	**0.912**
Precision	0.518	0.535	0.415	**0.592**
Recall	**0.797**	0.748	0.104	0.596
AUC PRC	**0.702**	0.691	0.201	0.602

Experiment 1 Confusion Matrix

Experiment 2 Confusion Matrix



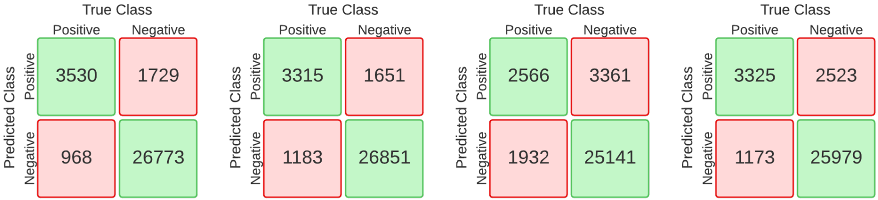
Experiment 3 Confusion Matrix



Experiment 4 Confusion Matrix

Results: Wall Rearing

	ResNet+LSTM	Inc.ResNet+LSTM	ResNet+TCN	Inc.ResNet+TCN
Binary Accuracy	**0.918**	0.914	0.840	0.888
Precision	**0.671**	0.667	0.433	0.569
Recall	**0.785**	0.737	0.570	0.739
AUC PRC	**0.792**	0.770	0.411	0.716



Experiment 1 Confusion Matrix | Experiment 2 Confusion Matrix | Experiment 3 Confusion Matrix | Experiment 4 Confusion Matrix

Results: Discussion

Groomings

	ResNet+LSTM	Inc.ResNet+LSTM	ResNet+TCN	Inc.ResNet+TCN
Binary Accuracy	0.902	0.909	**0.950**	0.942
Precision	0.412	0.426	**0.681**	0.577
Recall	**0.946**	0.876	0.537	0.635
AUC PRC	0.679	**0.690**	0.588	0.586

Mid Rearings

	ResNet+LSTM	Inc.ResNet+LSTM	ResNet+TCN	Inc.ResNet+TCN
Binary Accuracy	0.897	0.902	0.887	**0.912**
Precision	0.518	0.535	0.415	**0.592**
Recall	**0.797**	0.748	0.104	0.596
AUC PRC	**0.702**	0.691	0.201	0.602

Wall Rearings

	ResNet+LSTM	Inc.ResNet+LSTM	ResNet+TCN	Inc.ResNet+TCN
Binary Accuracy	**0.918**	0.914	0.840	0.888
Precision	**0.671**	0.667	0.433	0.569
Recall	**0.785**	0.737	0.570	0.739
AUC PRC	**0.792**	0.770	0.411	0.716

Conclusions

- Different architectures have been reviewed from the current literature.
- Main Goal: four different models to automatically tag the videos.
- Experimentation using LSTM and TCN networks.
- Each behavior has a different best model.
- Results show that the model that best suits and generalizes the problem is **ResNet + LSTM**

Conclusions

Future Work

- Try to use Transformers on this problem.
- Build the model together and perform a two stage training.
- Record the videos from different points of view.
- Try to use an ensemble method to combine the four models trained.
- Use the computational time as another metric.
- Increase the hyperparameter space to search and use other techniques.
- Test the models in public datasets to compare with other researchers.

Questions?