Leg movement detection for sleep micro-event analysis

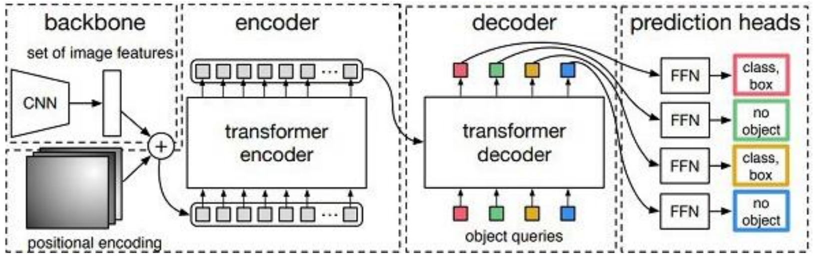
Siham kida (s210154), Jedrzej Konrad Kolbert (s184361), Matteo Sabadin (s212980), Muse Ali (s194615).
Technical University of Denmark

Introduction

Electromyography (EMG) is a diagnostic procedure to assess the health of muscles and the nerve cells that control them (motor neurons). EMG results can **reveal nerve dysfunction, muscle dysfunction** or problems with nerve-to-muscle signal transmission. In sleep medicine, clinicians interpret raw EMG signals, relying on shorter prototypical micro-architecture events which exhibit variable duration and shapes. Such events are traditionally **annotated by hand** by trained **sleep experts**, making the process **time consuming** and **subjective**. The aim of this project is to automate this procedure to make it **faster** and more **reliable**, by applying the **DE**tection **TR**ansformer framework to a set of recorded EMGs[3] related to the **leg movement** during sleep.

Model

DETR consists of a **convolutional backbone** followed by an **encoder-decoder Transformer** which can be trained end-to-end for object detection. First, an image is sent through a pre-trained convolutional backbone, the **ResNet-50**, that outputs a new **lower-resolution feature map**. This is then projected to the Transformer of DETR, using a nn.Conv2D layer. The flattened output is a tensor that represents the number of object queries, that is sent through the decoder. The decoder updates these embeddings through multiple attention layers. Next, two heads are added on top for object detection: a linear layer for **classifying** each object query into one of **the objects or "no object"**, and a MLP to predict bounding boxes for each query. The model is trained using a bipartite matching loss, that compares both the predicted classes and bounding boxes of each of the object queries. The parameters are optimized using the **matching algorithm** and standard cross-entropy for the annotations, L1 and IoU losses for the bounding boxes. In our first approach, the pairs of 1-dimensional timeseries sleep data were transformed into **spectrogram** representations, and then passed through an **additional CNN layer** to obtain the 3-dimensional input requested by the DETR model.



In the second approach, we directly used the spectrogram images for training on the original model, without using the additional CNN.

The bounding box loss function used to plot the results is :

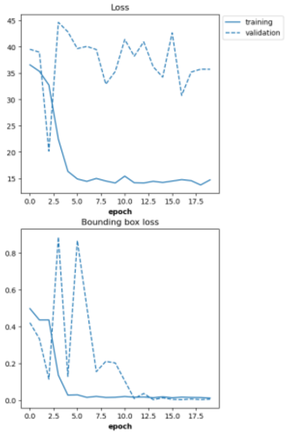
$$\mathcal{L}_{\text{box}}\left(\hat{b}_{\sigma(i)}, b_i\right) = \lambda_{\text{iou}}\,\mathcal{L}_{\text{iou}}\left(\hat{b}_{\sigma(i)}, b_i\right) + \lambda_{\text{L1}}\left\|b_i - \hat{b}_{\sigma(i)}\right\|_1$$

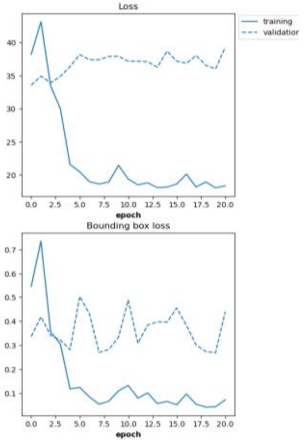
$\lambda_{iou}, \lambda_{L1}$ - hyperparameters
$\hat{b}_{\sigma(i)}$- predicted box
b_i- ground truth box

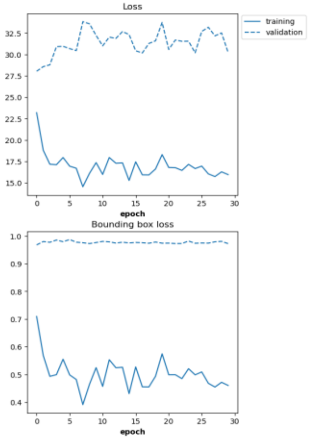
- Data: Images
- Batch size: 3
- Transformer layers: 6

- Data: Signals
- Batch size: 3
- Transformer layers: 6

- Data: Images
- Batch size: 8
- Transformer layers: 3

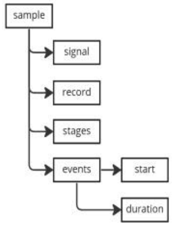


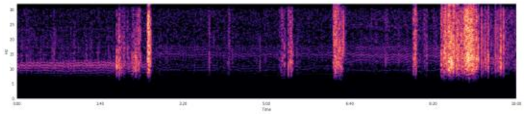


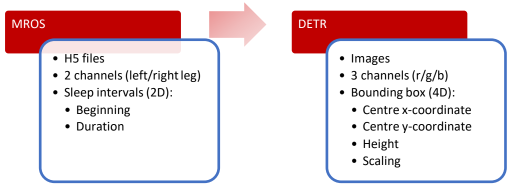


Data

The underlying data which this project was built upon were preprocessed recordings of .H5 files. Each sample of which structured as seen on the right figure. A sample contains **2 channels** (respectively for left and right leg) and **4801 timepoints**. Using a provided datamodule the .H5 files gets loaded and prepared for entering the model for training.

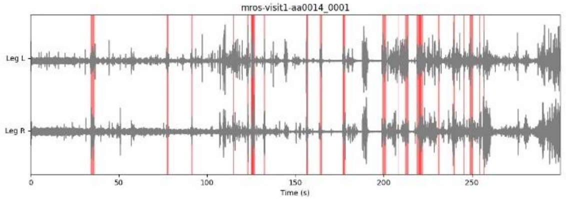




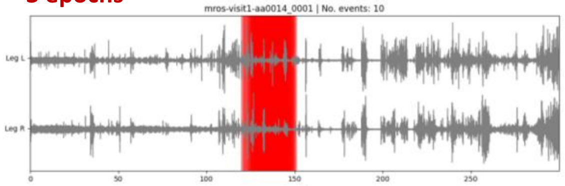


Results

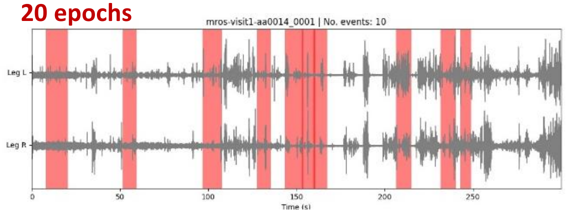
Ground truth



3 epochs



20 epochs



Conclusion

It is evident that our results does not provide high fidelity sleep event detection as our scoring of leg movement event shows not to be completely accurate. One of the biggest barriers for the training of model were **compute power**, and we determine that the model prediction would be precise by running **more epochs** and **deeper layers** of encoders and decoders. It would also be interesting to investigate **different backbones** to extract feature representation of the input samples. Moreover, adjustment of loss function from evaluation of **bounding box to "bounding interval"** could enhance learning rate and results quality, as well as including the information related to the **sleep phases** into the model.

References

[1] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In European conference on computer vision (pp. 213-229). Springer, Cham.

[2]Alexander Neergaard Olesen et al. "Deep transfer learning for improving single-EEG arousal detection".

[3] Blackwell, T., Yaffe, K., Ancoli-Israel, S., Redline, S., Ensrud, K.E., Stefanick, M.L., Laffan, A., Stone, K.L. and (2011), Associations Between Sleep Architecture and Sleep-Disordered Breathing and Cognition in Older Community-Dwelling Men: The Osteoporotic Fractures in Men Sleep Study. J Am Geriatr Soc, 59: 2217-2225. https://doi.org/10.1111/j.1532-5415.2011.03731.x