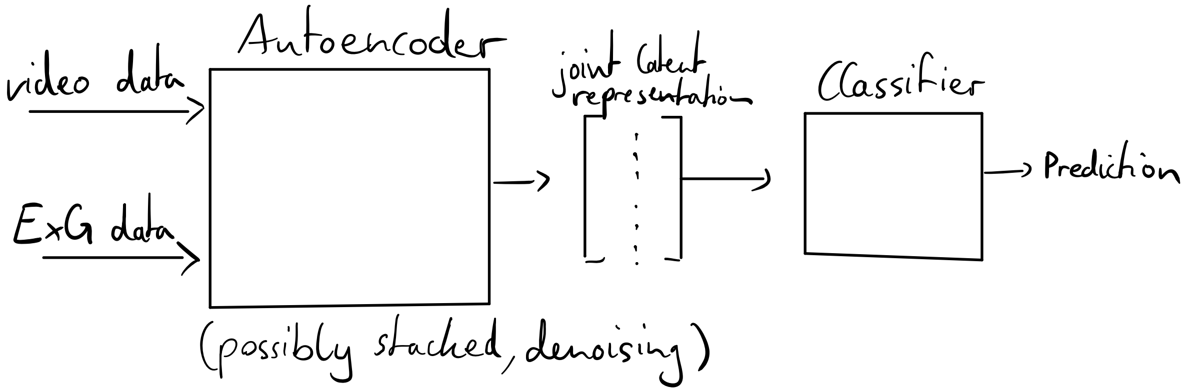
Summary (30.05.2022)

In this meeting, we talked about the papers I have read in preparation. These are presented below. Most of them were only interesting for a general overview of the topics (domain adaptation and co-learning), most of them were also outdated. Azade gave me a few more to read until next week (mentioned further down) and we talked about a general scheme we can apply to solve the problem of learning from video + ExG without needing the video data at the test phase. In general, we think it makes sense to use some type of autoencoder (possibly a denoising autoencoder?) to learn a joint representation of the data in latent space (to „link" video and ExG data, so to speak). A classifier might be used to classify the joint representation afterwards.



To-Do for next week:

We decided that it would make sense to test this idea on available data, to see if this is feasible at all. I will try to implement a simple autoencoder and classifier in the coming week(s). Additionally, I will try to find and read more recent papers on the topic (2017 or newer, hopefully).

Papers read:

- Sener, Ozan, et al. "Learning transferrable representations for unsupervised domain adaptation." *Advances in neural information processing systems* 29 (2016).
    - Short summary
        - Tries to learn a transformation between source examples (for us: video) and target examples (for us: ExG)
        - Tries to label unsupervised points (points in the target domain)
        - Examples used for source and target domain are MNIST and SVHN (both of which are datasets of images that contain numbers)
    - Good
        - Introduction of a deep learning framework
    - Bad

        - Assumes that the two domains are defined on the same space (not the case for us)
        - Assumes that the target domain is unsupervised (unlabelled) whereas the source domain is supervised (not the case for us)
- Sun, Shiliang. "A survey of multi-view machine learning." *Neural computing and applications* 23.7 (2013): 2031-2038.
    - Short summary
        - Contains explanations on concepts in multi-vew machine learning (only a small amount of which are interesting to us)
        - Example: Multi-view transfer learning
            - Area of machine learning that tries to „bridge a gap between source and target domain whose distributions can differ substantially" (for us: video data and ExG data)
    - Good
        - Gives an overview on multi-view machine learning
        - Mentions co-learning and multi-view transfer learning
    - Bad
        - The difference to domain concepts like domain adaptation or multimodal learning isn't that clear
        - A lot of concepts are presented that aren't useful to us
- Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." *IEEE Transactions on knowledge and data engineering* 22.10 (2009): 1345-1359.
    - Short summary
        - Gives an overview on transfer learning and its subclasses
    - Good
        - Contains a classification of transfer learning schemes in terms of distance between domains and tasks
            - Introduces „transductive transfer learning" as a subclass of transfer learning where source and target domain are „different but related" and source and target tasks are the same
            - Talks about domain adaptation as one way to do transductive transfer learning
        - Cites a lot of sources for every type of transfer learning
    - Bad
        - Old
        - Lots of unnecessary information
- Glorot, Xavier, Antoine Bordes, and Yoshua Bengio. "Domain adaptation for large-scale sentiment classification: A deep learning approach." *ICML*. 2011.

- Short summary
    - Tries to find an approach to model sentiment classifiers using domain adaptation
    - Uses Amazon reviews in different categories for domain adaptation (i.e. Toys, Software, Apparel etc.)
    - Tries to find a joint representation in latent space
    - Very similar to our ideas
  - Good
    - Uses an autoencoder (or rather, stacked autoencoders) to find a joint representation in latent space
    - Uses a classifier as a second step to classify the representation (they use an SVM)
  - Bad
    - Assumes that other domain(s) are unlabelled
    - Very old paper (2011)
- Chen, Minmin, Kilian Q. Weinberger, and John Blitzer. "Co-training for domain adaptation." *Advances in neural information processing systems* 24 (2011).
  - Short Summary
    - Adapt a model to a changing input domain
    - Adds examples from a target domain to the source domain during training (according to confidence levels)
  - Good
    - Seems like a very simple yet effective concept
    - Their results beat „normal" logistic regression for small amounts of labelled target data
  - Bad
    - Uses logistic regression for classification (not that useful for us, but the concept could be applied to other types of machine learning)
    - Old! (2011)
- Dai, Wenyuan, et al. "Translated learning: Transfer learning across different feature spaces." *Advances in neural information processing systems* 21 (2008).
  - Short summary
    - Tries to find a mapping between two separate domains
    - In this case between text and images
    - Calls this „translated learning"
  - Good
    - Describes a method for very different domains (text vs. images)
    - Simple concept (find a mapping between domains)

- Bad
  - Hard to apply the given building blocks to our problem (separating the data into four sets of data pairs)
  - No straight forward implementation
  - Very old (2008)

Papers Azade sent me (for next week):

- Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International Conference on Machine Learning*. PMLR, 2021.
- Sung, George, et al. "On-device Real-time Hand Gesture Recognition." *arXiv preprint arXiv:2111.00038* (2021).
- Guo, Wenzhong, Jianwen Wang, and Shiping Wang. "Deep multimodal representation learning: A survey." *IEEE Access 7* (2019): 63373-63394.