Машинное обучение для работы с видео

Общие идеи для работы с видео

- 3D свёртки по времени и пространству (минусы: много параметров, вычислительно дорого)
- Факторизация 3D свёрток на 2D для пространства и 1D для времени
- Two-Stream подход
- Inflated 3D CNN – использование предобученных на картинках 2D свёрток для начального приближения
- Использование разных типов attention

Задачи, на которые обратить внимание

- Video-to-Video
- Озвучивание видео

Статьи

- Серия обзоров методов работы с видео (https://towardsdatascience.com/deep-learning-on-video-part-one-the-early-days-8a3632ed47d4)
- Inflated 3D CNN (https://arxiv.org/pdf/1705.07750.pdf)

Предобученные модели

- https://github.com/facebookresearch/SlowFast

Speech Prediction in Silent Videos using Variational Autoencoders (https://arxiv.org/abs/2011.07340)

Обзор статьи

- Решается задача озвучивания видео с применением вариационного автоэнкодера
- Во время обучения на вход подается последовательность кадров и звуковой сигнал
- Сначала с помощью энкодеров получают эмбеддинги независимо для каждого аудио/видео фрейма
- Далее применяют LSTM и полносвязный слой для получения среднего и дисперсии вариационных распределений q
- Оптимизируют ELBO (evidence lower bound)

$$\mathcal{L}(\theta, \phi; \mathrm{x}) = \sum_{t=1}^{N} \mathbb{E}_{q_{\phi_a}(z|a_t)}[\lambda \log p_{\theta_a}(a_t|z)] - \beta KL[q_{\phi_a}(z|a_t)||q_{\phi_f}(z|f_t)]$$

Комментарий

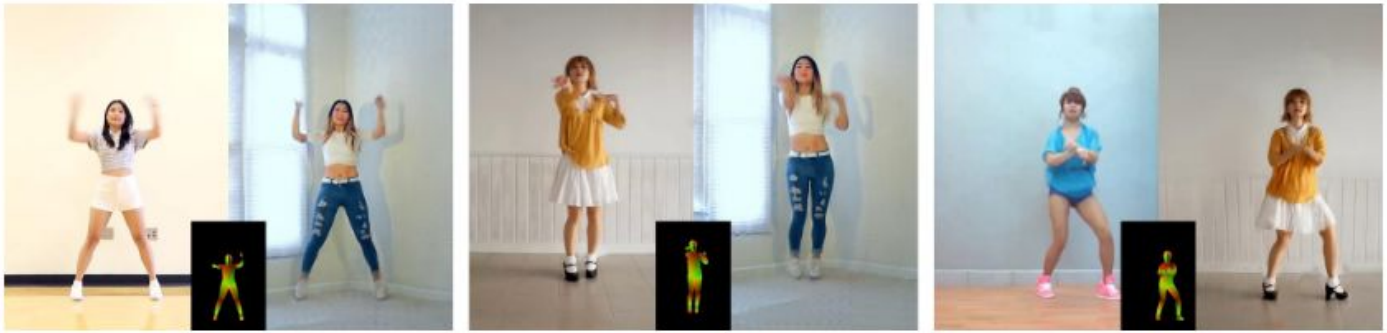
- Чтобы решать нашу задачу, нужно будет заменить первичную обработку звукового сигнала (перевод в спектрограмму и audio encoder), а также последующее восстановление сигнала из спектрограммы (audio decoder, Griffin-Lim reconstruction) на соответствующие компоненты для обработки fMRI
- Чтобы не обучать модель отдельно для каждого участника, необходимо добавить закодированную информацию о человеке (conditional VAE)

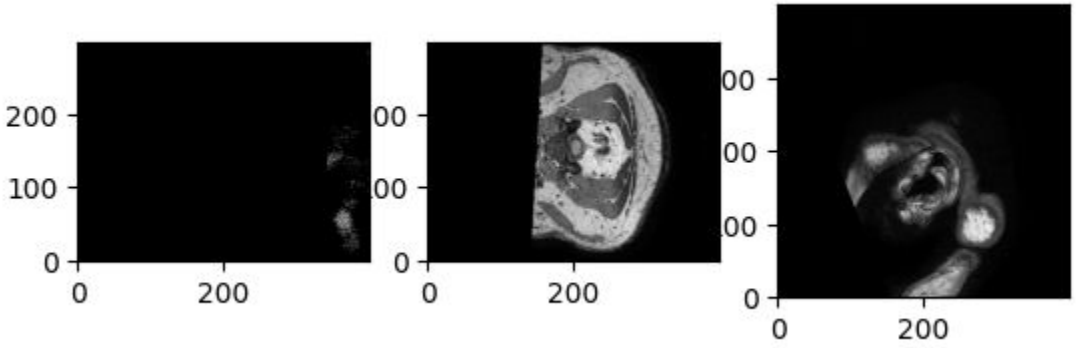
Video-to-Video Synthesis

(https://arxiv.org/abs/1808.06601)

Обзор статьи

- Решается задача генерации нового видео по исходному
- Модель conditional GAN
- Markov assumption
- Используются два дискриминатора: для отдельных кадров и для видеоряда
- Функция потерь содержит компоненту, отвечающую за согласованность генерируемого сигнала




200
100
0
0
200
00
00
0
0
200
00
00
00
0
0
200

fMRI: preprocessing, classification and pattern recognition

Обзор статьи

- Исследуется задача классификации (например, психических расстройств) по снимкам FMRI
- Проблема – fMRI слишком шумный сигнал
- Шум возникает от движения головы, биения сердца, температурного фона и т.д.
- Рассматриваются несколько подходов шумоподавления
- Также рассматриваются новые Feature Extraction подходы, основанные на топологическом анализе данных
- Предлагается новый pipeline для решения задач неврологии
- Показывается эффективность на задачах определения эпилепсии и депрессии

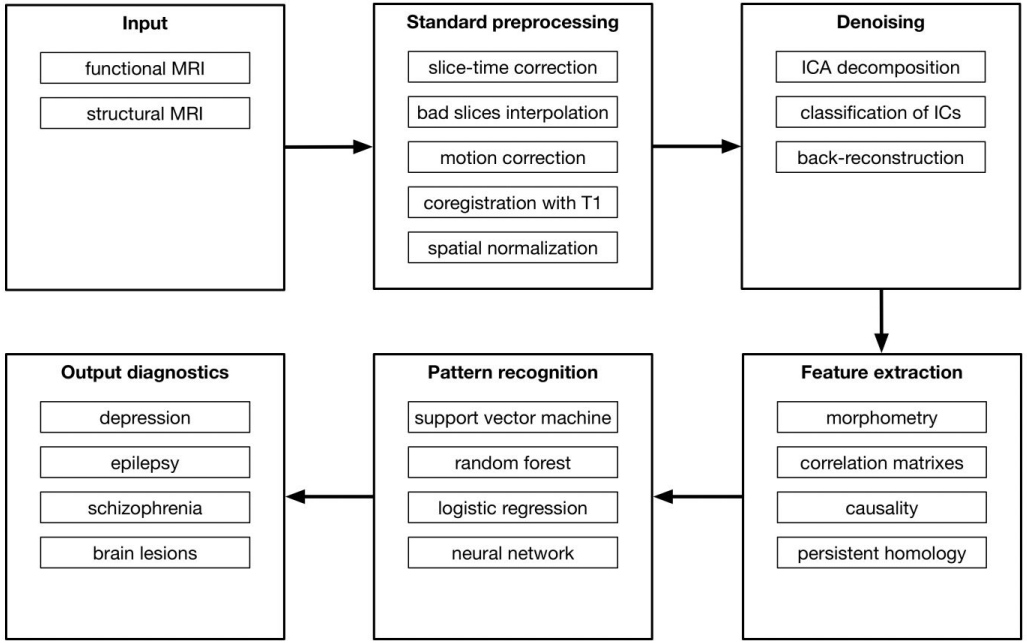


Figure 1: The scheme of the proposed noise-aware fMRI processing pipeline.

https://arxiv.org/pdf/1804.10167v1.pdf