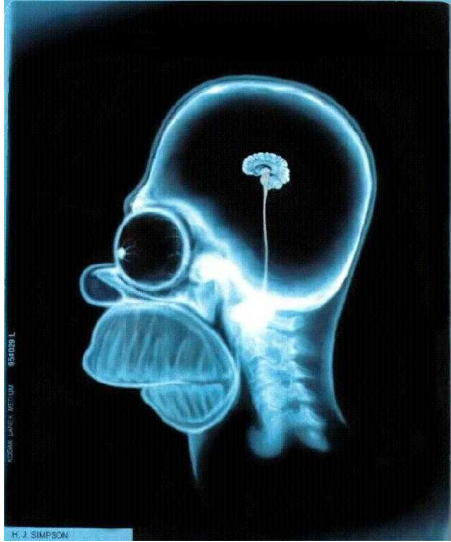
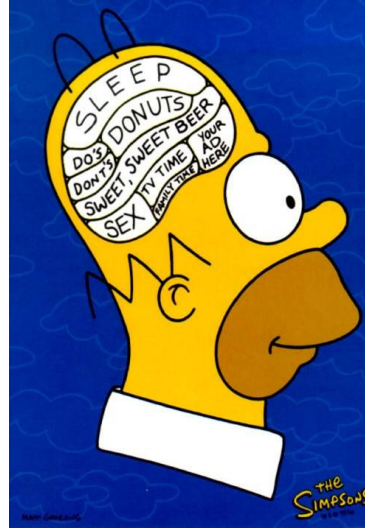


## MRI vs. fMRI

MRI (Magnetic Resonance Imaging) studies brain anatomy.



Functional MRI (fMRI) studies brain function.

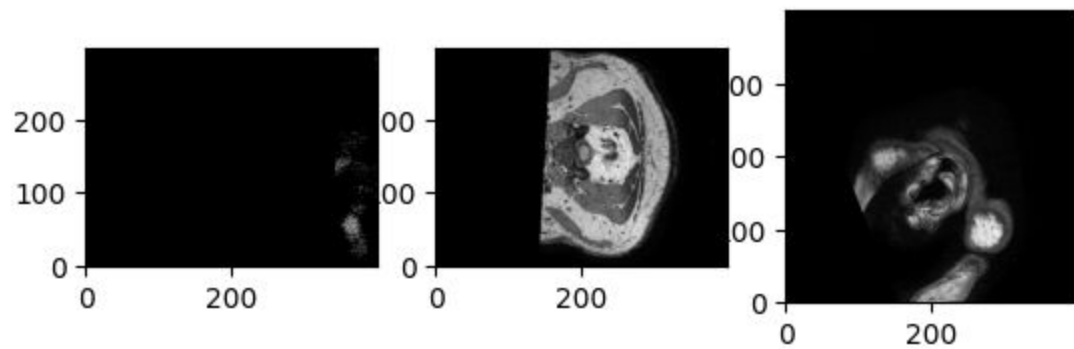


<https://lukas-snoek.com/NI-edu/index.html>

[https://andysbrainbook.readthedocs.io/en/latest/fMRI\\_Short\\_Course/fMRI\\_Intro.html](https://andysbrainbook.readthedocs.io/en/latest/fMRI_Short_Course/fMRI_Intro.html)

<https://fmriprep.org/en/stable/#>

[https://nipy.org/nibabel/coordinate\\_systems.html#introducing-someone](https://nipy.org/nibabel/coordinate_systems.html#introducing-someone)



Data type	Task/Application	Method	Year
fMRI	Survey	-	2001
fMRI	Survey	-	1994
fMRI	Survey	-	2010
fMRI	Survey	-	2008
fMRI	Survey	-	2012
fMRI	Classification	Machine learning	2009
fMRI	fMRI data analysis	SVM	2007
rs-fMRI	fMRI analysis	ICA	2016
rs-fMRI	Classification	Sparse dictionary learning	2012
t-fMRI	Feature engineering	Sparse dictionary learning	2009
t-fMRI	Classification	Machine learning	2014
fMRI	Brain response analysis	Machine learning	2008
fMRI	Classification	Hybrid Machine learning	2010
rs-fMRI	Segmentation	Graph analysis	2014
rs-fMRI	FC extraction	Matrix factorization	2013

rs-fMRI	Classification	Deep learning	2019
rs-fMRI	Classification	RNN	2018
rs-fMRI	Classification	LSTM	2017
rs-fMRI	Prediction	CNN	2018
rs-fMRI	Classification	3D CNN	2017
rs-fMRI	Pattern extraction	RBM	2018
t-fMRI	Pattern extraction	CNN	2018
t-fMRI	Feature extraction	CNN	2018
t-fMRI	Pattern extraction	CNN	2017
t-fMRI	Classification	Deep learning	2018
t-fMRI	Classification	3D CNN	2018
t-fMRI	Brain decoding	LSTM	2018
t-fMRI	Pattern extraction	RNN	2016
t-fMRI	Pattern extraction	LSTM	2018
t-fMRI	Disease analysis	RNN	2017
rs-fMRI	Survey	-	2010
rs-fMRI	Survey	-	2008
rs-fMRI	Survey	-	2007
rs-fMRI	FC extraction	ICA	2011
rs-fMRI	FC extraction	Machine learning	2014
rs-fMRI	FC extraction	Machine learning	2013
rs-fMRI	FC extraction	ICA	2014

Name of study	Subjects & sample size	Machine learning method/classifier	Accuracy achieved
Qureshi et al. (2017)	SCZ = 72, HC = 72	ELM	99.3%
Johnston, Tolomeo, et al. (2015)	HC = 21, MDD = 19	SVM with leave one out cross validation	97.00%
Wang, Ren, and Zhang (2017)	HC = 29, MDD = 31	SVM with leave one out cross validation	95.00%
Khazaee, A (2017)	HC = 45, AD = 34	Naïve bayes	93.29 %
Gao et al. (2017)	BD = 37, MDD = 37	SVM with 10-fold cross validation	93.00%
Matsubara et al. (2019)	SCZ = 54, HC = 67	SVM with 10-fold cross validation	92.00%
Arbabshirani et al. (2014)	SCZ = 195, HC = 175	Functional connectivity and SVM	88.21%
Li et al. (2017)	BD = 22, MDD = 22	SVM with leave one out cross validation	86.00%
Johnston, Steele, et al. (2015)	HC = 21, MDD = 25	SVM with leave one out cross validation	85.00%
Chen et al. (2017)	HC = 31, ASD = 22	SVM	83.33%
Yoshida et al. (2017)	HC = 65, MDD = 58	PLSR with leave one out cross validation	80.00%
Rubin-Falcone et al. (2018)	BD = 26, MDD = 26	SVM with leave-two out cross validation	75.00%
Khazaee A (2017)	45 HC, AD = 34	SVM with radial base function	71.95%
Sankar et al. (2016)	HC = 20, MDD = 23	SVM with 5-fold cross validation	70.00%
Deng et al. (2018)	BD = 31, MDD = 36	SVM with leave one out cross validation	68.30%

Abbreviations: AD, Alzheimer's disease; ASD, autism spectrum disorder; BD, bipolar disorder; ELM, extreme learning machine; HC, health controls; MDD, major depressive disorder; SCZ, Schizophrenia; SVM, support vector machine.

**TABLE 8** Deep learning studies with highest accuracies in healthy controls versus Alzheimer's disease

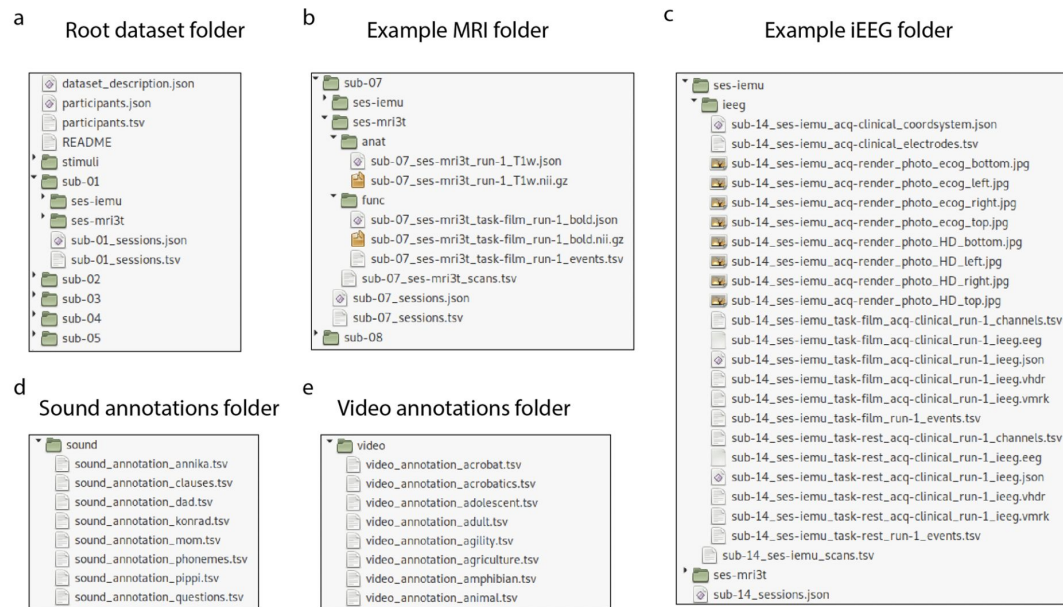
Name of study	Dataset used	Deep learning architecture	Accuracy achieved
Basaia (2019)	ADNI	Convolutional neural network	99.2
Suk (2015)	ADNI	Stacked Autoencoder	98.8
Hosseini-Asl et al. (2016)	ADNI	Convolutional neural network	97.6
Sarraf and Tofighi (2016)	ADNI	Convolutional neural network	96.9
Suk and Shen (2013)	ADNI	Stacked Autoencoder	95.9
Suk (2015)	ADNI	Deep weighted subclass-based sparse multi-task learning	95.1
Suk (2014)	ADNI	Deep Boltzman Machine	94.9
Gupta et al. (2013)	ADNI	Sparse auto-encoder with convolutional neural network	94.7
Li et al. (2014)	ADNI	Deep multilayer perceptron	91.4
Suk (2017)	ADNI	Multiple sparse regression with convolutional neural network	91.02
Liu, Liu, Cai, Che, et al. (2014)	ADNI	Stacked Autoencoder	87.8



# Обзор статьи

## Некоторые характеристики датасета

- 30 участников с данными FMRI
- 13 чередующихся эпизодов по 30 секунд (речь / музыка)
- Всё переведено на голландский
- аудио и видеодорожка аннотированы (время начала и конца появления каждого объекта)
- Снимки FMRI в формате BIDS
- Папка видео → 135 tsv файлов (для 129 объектов, и 6 персонажей)



**Fig. 2** Overview of data records. (a–e) Structure of folders and files in the dataset with example (f) MRI and iEEG folders from one subject.

# Speech Prediction in Silent Videos using Variational Autoencoders

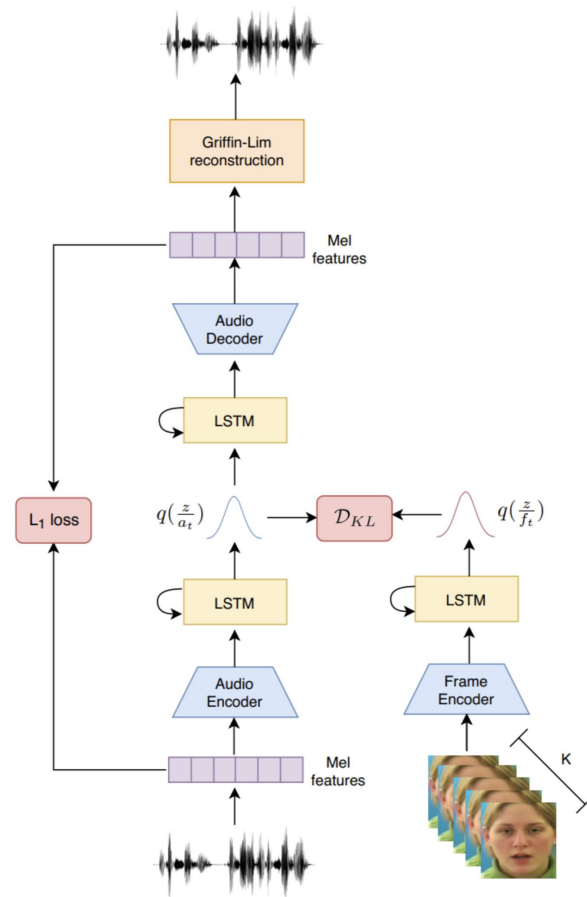
## Обзор статьи

- Решается задача озвучивания видео с применением вариационного автоэнкодера
- Во время обучения на вход подается последовательность кадров и звуковой сигнал
- Сначала с помощью энкодеров получают эмбединги 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)

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

## 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
- В исходном виде данную модель придется обучать отдельно для каждого участника эксперимента



# 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 эмбединги 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)

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

## Комментарий

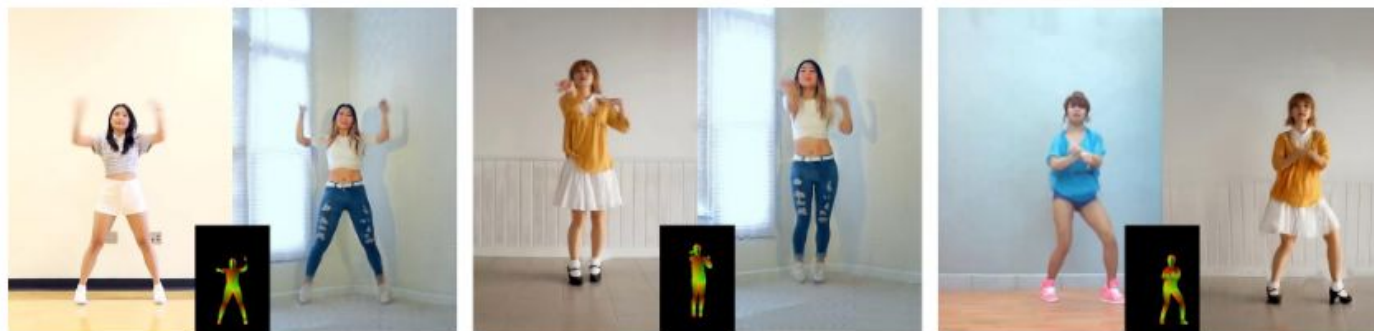
- Чтобы решать нашу задачу, нужно будет заменить первичную обработку звукового сигнала (перевод в спектрограмму и audio encoder), а также последующее восстановление сигнала из спектрограммы (audio decoder, Griffin-Lim reconstruction) на соответствующие компоненты для обработки fMRI
- Чтобы не обучать 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



# 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

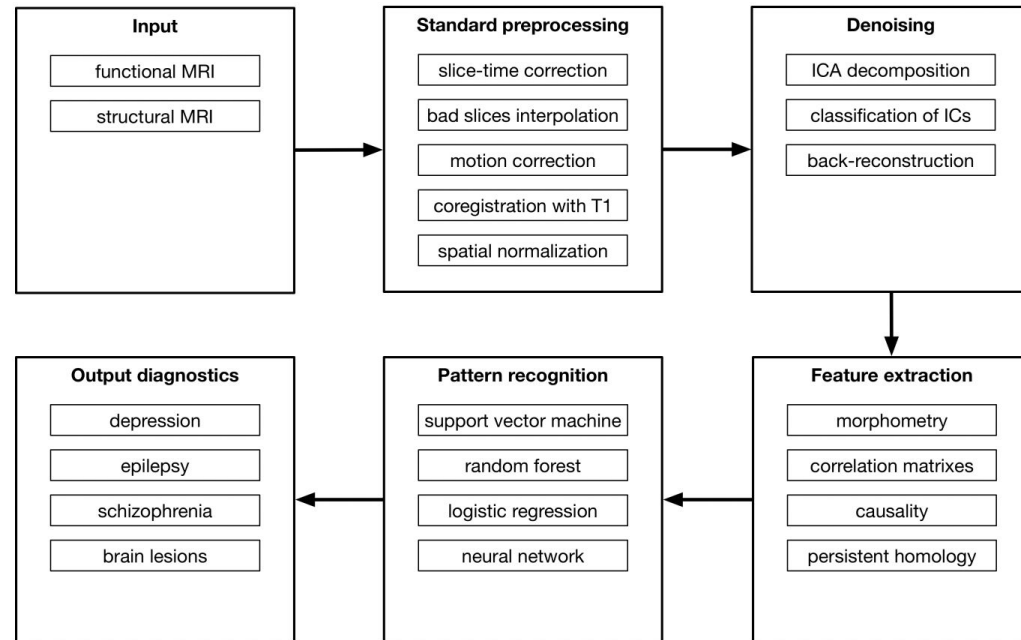


Figure 1: The scheme of the proposed noise-aware fMRI processing pipeline.