

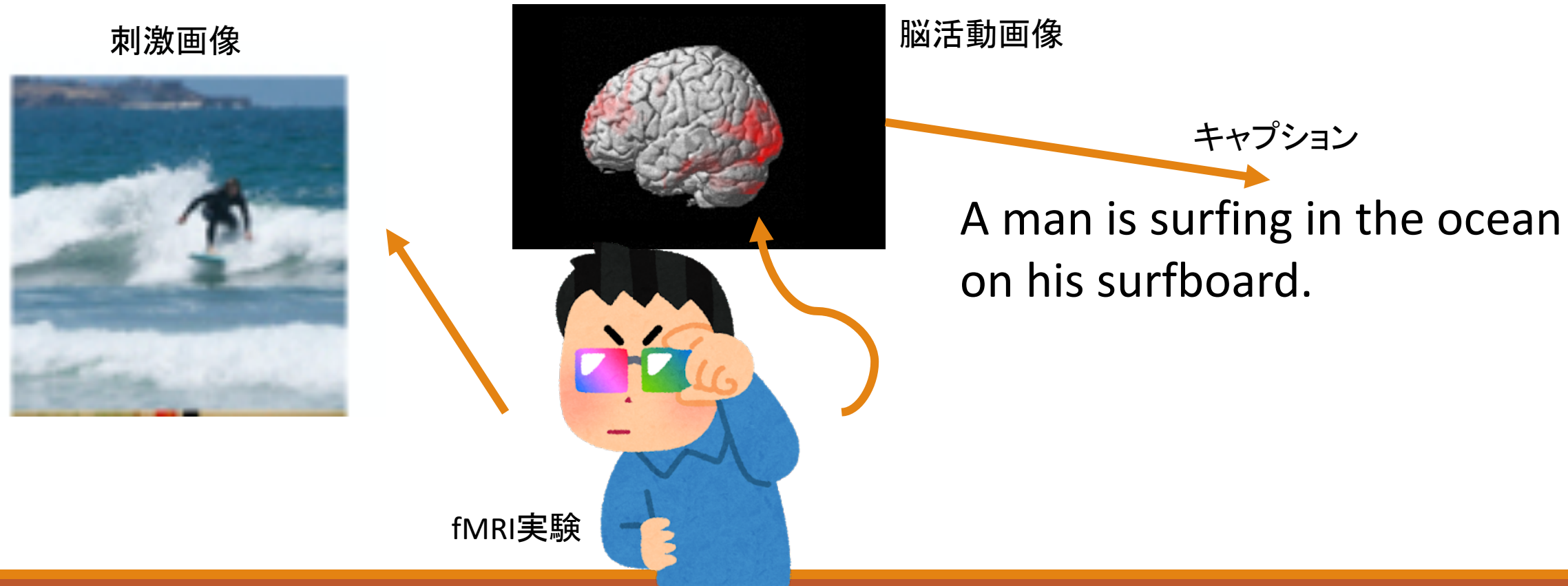
論文輪読2
深層学習による
画像刺激時のfMRI脳活動データからの文生成

目次

1. 何をやったのか？
2. 何を使っているのか？
3. 何がわかった？
4. 次は何を読めばいいんだろうか？-今回はやらない

何をやったのか？

画像刺激時のfMRI脳活動データを使って、
その刺激の画像のキャプションを生成する。



何を使っているのか？

1. 刺激画像->画像特徴量->キャプション生成
2. 脳活動データ->画像特徴量
3. 脳活動データ->画像特徴量->キャプション生成

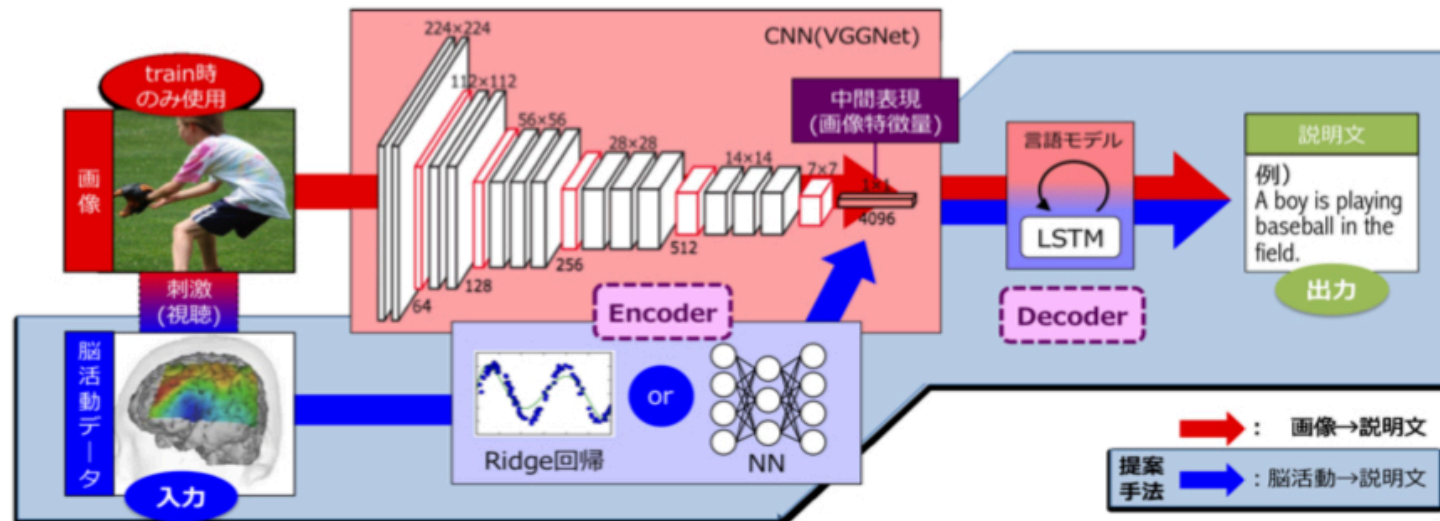


図 1: 本研究の概要

パラメータ





表 1: 実験設定 (詳細)

	(A) 画像→画像特徴量→説明文モデル	(B) 脳活動データ→画像特徴量モデル		
		Ridge 回帰	3 層 NN	5 層 DNN
データセット	Microsoft COCO	動画刺激による脳活動データ		
学習量	414,113 sample×100 epoch	4,500 sample×1,000 epoch		
アルゴリズム	Adam	Ridge regression	stochastic gradient descent	
学習に関する ハイパーパラ メータ	a=0.001, b1=0.9, b2=0.999, eps=1e-8 勾配閾値: 1 L2 正則化項: 0.005	L2 正則化項: 0.5	学習率: 0.01 勾配閾値: 1 L2 正則化項: 0.005	
学習する パラメータ の初期値	word embedding: word2vec VGGNet: 事前学習済み・学習せず それ以外: 標準正規分布乱数	標準正規分布乱数	標準正規分布乱数	教師なし脳活動データを用いた AutoEncoder による事前学習 (7,540sample×200epoch)
層ユニット数	各層 512	65,665 - 4,096	65,665 - 8,000 - 4,096	65,665 - 7,500 - 6,500 - 5,500 - 4,096
語彙	頻出語 3,469 語	—		
誤差関数	交差エントロピー	平均二乗誤差		

何がわかった？結論

脳活動画像から生成した文と画像から生成した文



表 3: 実験 (C-1) : 被験者が見ていた画像, その時の脳活動から生成した説明文 3 通り, 画像から生成した説明文の例

	Ridge 回帰	3 層 NN	5 層 DNN	Image → Caption Model
train data		A group of people walking down the street.	A group of people walking down the street.	A group of people walking down the street.
		A pair of scissors sitting on the ground.	A close up of an orange and white clock.	A pair of scissors sitting on the ground.
test data		A bench sitting in the middle of some grass.	A fire hydrant sitting on the side of an empty street.	A herd of sheep standing in the grass.
		A bird sitting in top of an orange tree.	A fire hydrant sitting on the side of an empty street.	A bird is perched on top of the tree branch.

画像からのキャプションと同じものになるものをあつた。

何がわかった？結論

表 4: 実験 (C-2) : 被験者が見ていた画像と、その時の脳活動から生成した説明文 6 通り

	3,538 voxels (c.c. > 0.2)	5,961 voxels (c.c. > 0.15)	9,923 voxels (c.c. > 0.1)	21,437 voxels (c.c. > 0.05)	65,665 voxels (all cortex)	89,206 voxels (+subcortex)
	A young man is doing tricks on his skateboard.	A man is playing tennis on the court.	A young man is playing tennis on the court.	A man is playing tennis on the court.	A man is playing tennis on the court with his racket.	A man is playing tennis on the court with his racket.
	A man sitting on the ground with an umbrella.	A polar bear is standing in the water.	A dog laying on the ground next to an orange frisbee.	A dog laying on the ground next to an orange frisbee.	A black and white dog laying on the ground.	A dog is sitting on the floor in front of an open door.

テニスをしている男の人が表現できている。

犬のことは書いているが、内容は間違っている。

何がわかった？

脳画像->画像特徴量では、3層ニューラルネットワークが一番いい

深層学習を用いることで、刺激を受ける脳活動データの自然言語分表現への変換を実現した。

学習に用いる脳部位を限定して学習したモデルによる文生成を比較することで、人間の脳における画像認識処理が脳全体で行われていることを示唆する結果を得た。

論文の読み方 & まとめ方

