● 画像キャプションの様々な用途・性質

<視覚障がい者向け>
People wearing masks are sitting in the train.

<ニュース記事>
Commuters in Tokyo in April.*

● 画像キャプションの用途に応じて，多様なキャプションを生成

■ 描写度合いを踏まえた画像キャプショニング

描写度合い
0.7 → A black bear is walking in the trees.
0.4 → An animal in the forest.

*https://www.nytimes.com/2020/06/06/world/asia/japan-coronavirus-masks.html

心像性（imageability）[1]

● 「心的イメージの喚起しやすさ」を表す単語概念
● 文に拡張し，「文の心像性」を提案・推定 [2]

	“bear”		“animal”
心像性	0.8	>	0.4

➡ 画像キャプショニングの描写度合いとして使用

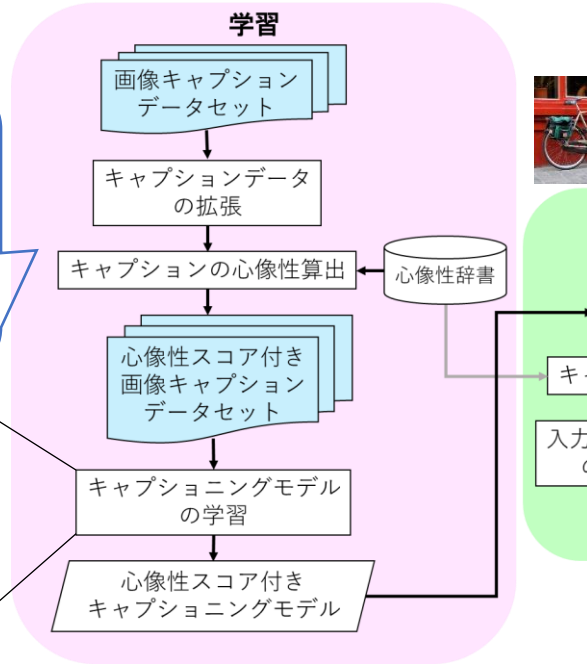
[1] Paivio et al., “Concreteness, imagery, and meaningfulness values for 925 nouns.,” J. Exp. Psychol, 1968.
[2] 梅村ら，“画像キャプションの質的評価に向けた文の心像性推定手法の検討”，NLP第25回年大，2019.

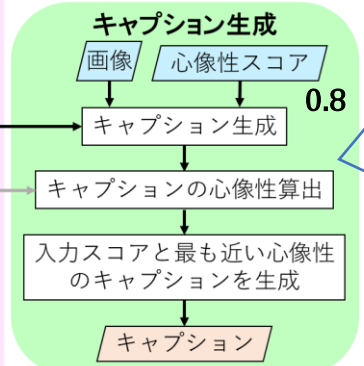
提案手法の処理手順

Cap1. Two brown horses in a pasture are eating the grass. → 0.85
Cap2. Two brown mammals in a pasture are eating the grass. → 0.68
Cap3. Two brown horses in a field are eating the grass. → 0.77
⋮

- x_t：単語特徴ベクトル
- I_t：アテンション付き画像特徴ベクトル
- IA：心像性特徴ベクトル

$$\begin{cases} \text{隠れ状態} h_t = \mathrm{LSTM}(\mathrm{concat}(x_t, I_t, IA)) \\ \text{次の単語} c_{t+1} = \mathrm{softmax}(h_t) \end{cases}$$

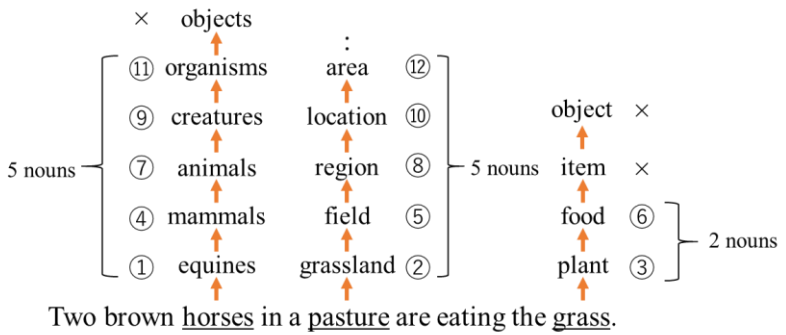


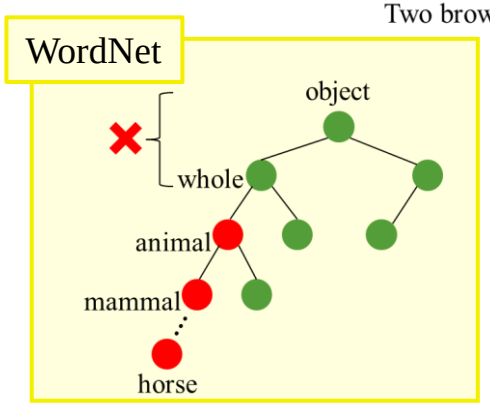


CapA. A dog sitting in front of a red door. → 0.59
CapB. A brown and white dog sitting on a leash. → 0.72
CapC. A brown and white dog laying next to a bike. → 0.77
CapD. A brown and white dog standing next to a red container. → 0.81
CapE. A white dog standing on the ground. → 0.63

A brown and white dog standing next to a red container.

キャプションデータの拡張





- MS COCO[3]データセット
  - キャプション5文 / 画像1枚
- 各キャプション中の各名詞を上位語に入れ替える
  - WordNet[4]の木構造を参照
  - 最大5階層上の語まで置換

[3] Lin et al., “Microsoft COCO Common Objects in Context.”, ECCV, 2014.
[4] Miller., “WordNet: A lexical database for English.”, Commun. ACM, 1995.

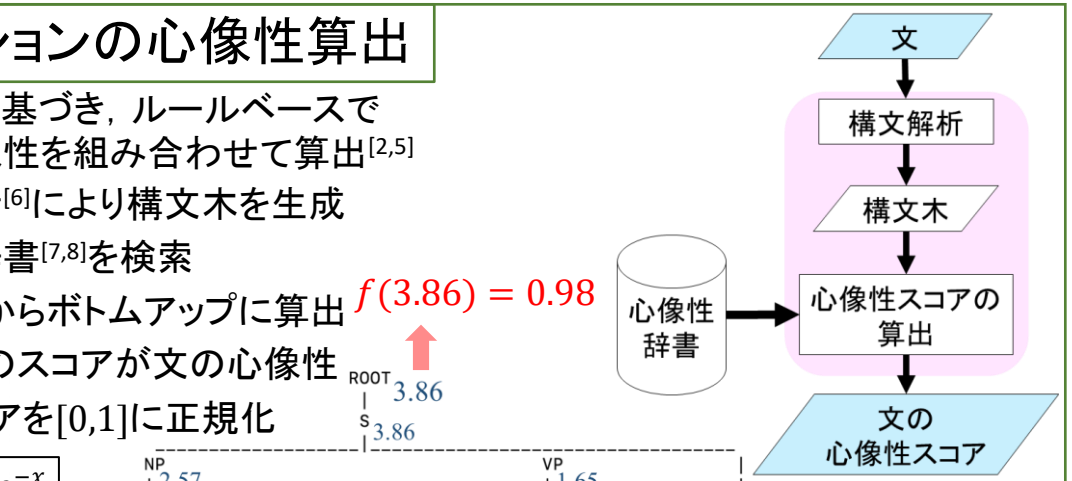
キャプションの心像性算出

- 構文木に基づき，ルールベースで単語心像性を組み合わせて算出[2,5]

1. 構文解析[6]により構文木を生成
2. 心像性辞書[7,8]を検索
3. 葉ノードからボトムアップに算出
4. 根ノードのスコアが文の心像性
5. そのスコアを[0,1]に正規化

$f(3.86) = 0.98$

$f(x) = 1 - e^{-x}$

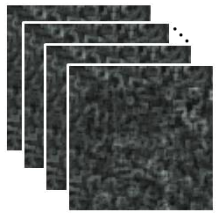


[5] 梅村ら，“心像性に基づく画像キャプショニングの検討”，信学技報2019-69，2020.
[6] Manning et al. , “The Stanford CoreNLP natural language processing toolkit”, ACL, 2014.
[7] Scott et al., “The Glasgow Norms: Ratings of 5,500 Words on Nine Scales.”, Behar. Res. Meth, 2018.
[8] Ljubešić et al., “Predicting concreteness and imageability of words within and across languages via word embeddings.”, Workshop on RL for NLP, 2018.

キャプショニングモデルの学習

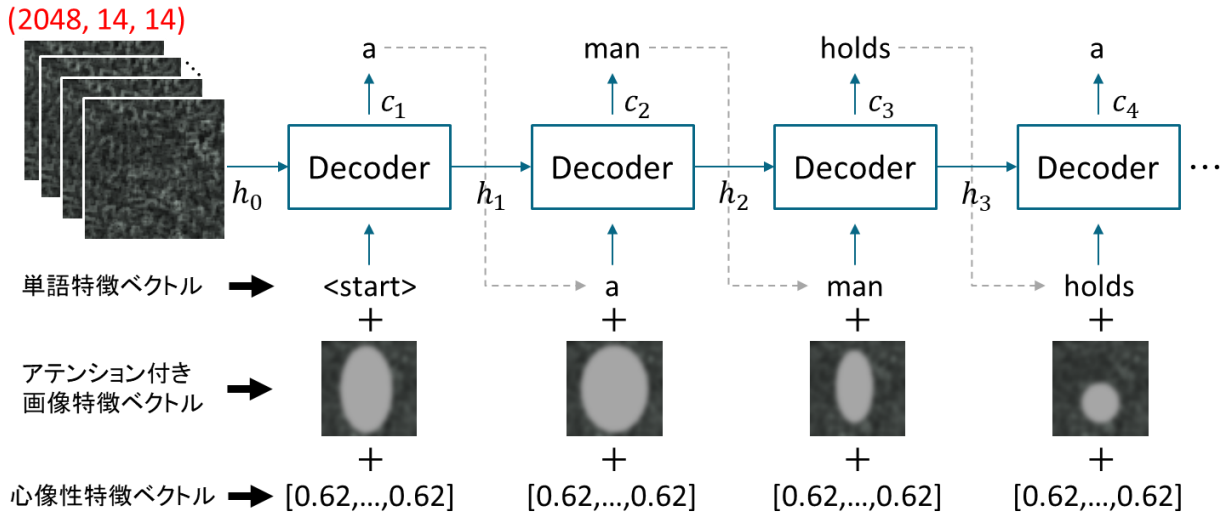
$(3, H, W)$

Encoder



$(2048, 14, 14)$

- ImageNetで学習済みのResNetを用いて画像特徴を抽出
- 画像特徴と心像性特徴をLSTMに入力し，1語ずつ生成
- 生成した単語ベクトルc_tと正解キャプション中の単語ベクトルw_tとの間のクロスエントロピーロスの最適化による，モデルの学習



実験設定

- 各画像につき20文のキャプションを使用して学習
  - 20文のサンプリング方法
    - ソートなし：データ拡張時の生成順に選択
    - ソートあり：キャプションを心像性順にソートし，最上位と最下位のキャプションを交互に選択
- 0.1, 0.2, … , 0.9の9種類の心像性を入力し，キャプション生成
  - low: [0.1, 0.3], mid: [0.4, 0.6], high: [0.7, 0.9]
- 比較手法
  - 候補キャプションを複数生成せず，最高精度のものを1つだけ生成

生成キャプションの心像性に関する分析

- 評価指標
  - 生成キャプションの異なり数
  - 生成キャプションの心像性の範囲（最大値－最小値）
  - 入力した心像性を真値とする平均二乗誤差（MSE）
  - 入力した心像性を真値とする平均平方二乗誤差（RMSE）

サンプリング方法	手法	異なり数	範囲	MSE			RMSE		
				low	mid	high	low	mid	high
ソートなし	提案	**4.68**	**0.083**	**0.405**	**0.118**	**0.011**	**0.632**	**0.334**	**0.098**
	比較	3.50	0.070	0.434	0.131	0.015	0.655	0.354	0.117
ソートあり	提案	**4.63**	**0.182**	**0.338**	**0.089**	**0.014**	**0.573**	**0.276**	**0.107**
	比較	3.26	0.164	0.378	0.103	0.022	0.607	0.300	0.142

画像キャプションの自動評価指標による評価

- 正解データ
  - 拡張後のキャプションデータ（20文/img）

サンプリング方法	手法	BLEU-4			CIDEr			ROUGE			METEOR			SPICE		
		low	mid	high	low	mid	high	low	mid	high	low	mid	high	low	mid	high
ソートなし	提案	0.270	0.268	0.263	0.675	0.677	0.676	0.502	0.501	0.501	0.233	0.236	0.240	0.088	0.090	0.092
	比較	**0.275**	**0.278**	**0.277**	**0.706**	**0.708**	**0.704**	**0.506**	**0.506**	**0.506**	**0.238**	**0.241**	**0.241**	0.089	**0.092**	0.092
ソートあり	提案	0.247	0.269	0.258	0.587	0.502	0.643	0.489	0.502	0.499	0.225	0.233	0.237	0.086	0.089	0.092
	比較	**0.251**	**0.274**	**0.276**	**0.607**	**0.651**	**0.652**	**0.492**	**0.505**	**0.506**	**0.228**	**0.236**	0.236	**0.087**	**0.092**	0.091

心像性スコア	生成キャプション
0.6	A placental is laying on a keyboard on a desk.
0.7	A vertebrate is laying on a keyboard on a desk.
0.8	A feline is laying on a keyboard on a desk.
0.9	A cat is laying on a computer keyboard.

心像性スコア	生成キャプション
0.6	A white and blue medium sitting on a runway.
0.7	A white and blue medium on a runway.
0.8	A small white and blue craft on a runway.
0.9	A small craft sitting on top of an airport tarmac.