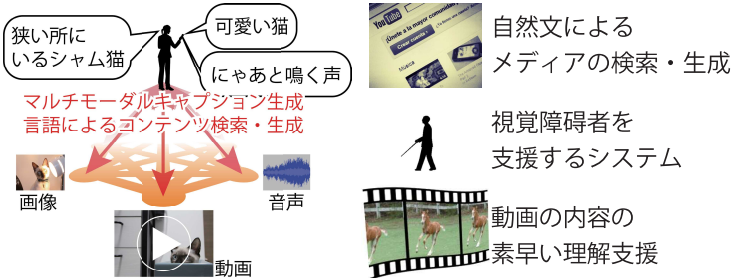


多様なデータへのキャプションを自動で生成する技術の創出 牛久祥孝 (OMRON SINIC X Corp.)

“視覚と言葉をつなげる技術”

未来ビジョン | 自然言語によるコンテンツ消費

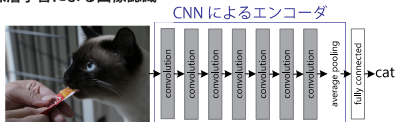
- 言語をクエリとしたメディアの検索・生成
- 視聴に時間のかかるメディアをキャプション生成して表示



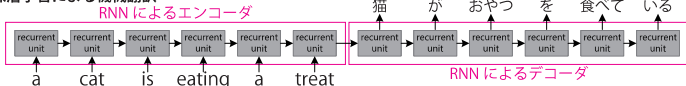
キャプション生成

- 言語によるメディア検索・生成にも必須
- キャプション生成の知見がメディア検索・生成に利用されている
- まずは多様なデータへのキャプション生成を実現したい

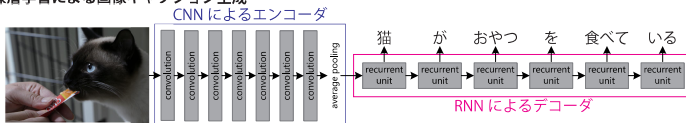
深層学習による画像認識



深層学習による機械翻訳

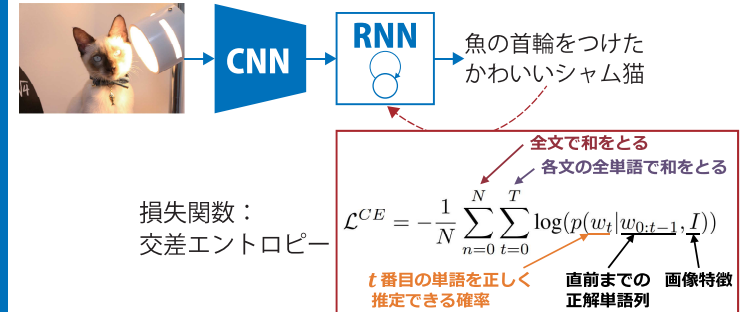


深層学習による画像キャプション生成



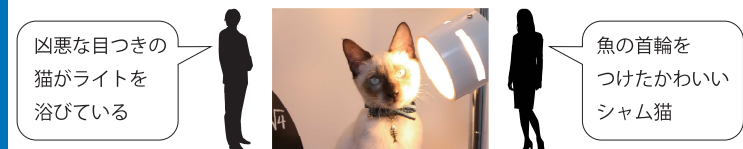
従来のキャプション生成手法では...

- 大多数が同意する客観的事実で
- 画像（動画）に大きく写ったものの抽象的な記述を
- 大量の画像（動画）とキャプションのペアデータから学習



未来ビジョン達成のために必須+未達の3要求機能

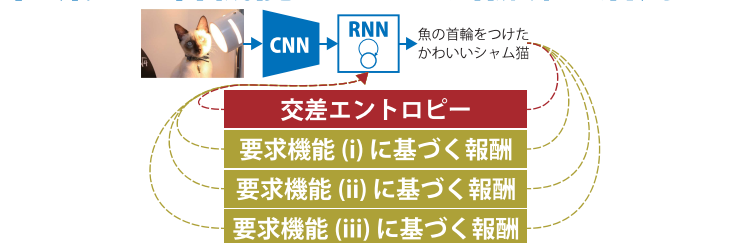
- 個人の属性や傾向への対応
- 詳細な表現への対応



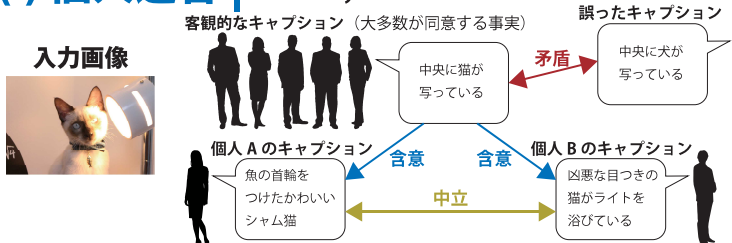
- 教師キャプションがないデータへの対応



本研究：各機能に基づく報酬の活用



(i) 個人適合 | Visually Grounded Entailment



(ii) 詳細な表現 | Referring Expression



Human-friendly Referring Expression Generation

- 人がすばやく見分けられる差分に注目した記述生成の提案
- ランク学習と視覚的顕著性+アテンションを活用した生成モデル



Human-friendly Referring Expression のためのデータセット収集

	Train	Val	Test
# images	23,950	2,395	2,405
# created instances	65,205	6,563	6,504
# referring expression	177,763	17,766	17,646

	Acc	Consistency	F1
Human in orange shirt	100%	2.49x10 ⁻²	1
A man in a black shirt next to a building	100%	4.23x10 ⁻²	1
A woman wearing a brown shirt and black shorts standing in the shade	100%	3.28x10 ⁻²	1
A man in a black shirt and grey pants walking on the sidewalk	100%	3.34x10 ⁻²	1
A man in an orange shirt	100%	4.37x10 ⁻²	1
A man in a black shirt and blue pants walking on the sidewalk	80%	3.32x10 ⁻²	1
A man in a black shirt and blue pants walking on the sidewalk	100%	6.85x10 ⁻²	1
A man in a black shirt and blue pants walking on the sidewalk	100%	6.85x10 ⁻²	1

今後の TODO | (iii) 教師キャプションなし

- すべてのデータに大規模キャプションを付与するのは困難
- 例えば ... 動画データに対して時間上に密なキャプションデータセットの作成は難しい
- シャム猫がゴミ箱から顔を出している
- 他の複数データセットの利用・ドメイン適応の活用