

視覚に基づく言い換えの分野開拓 (視覚に基づく言い換えのセマンティック類型)




チョ シンキ (CHU, Chenhui)
京都大学 特定准教授



1. 言語・視覚理解における言い換え認識の重要性

- 視覚的質問応答 [Wu+ 2017]



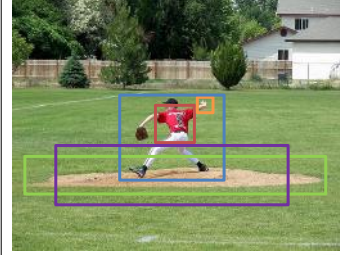
ユーザ1: what is **the man** doing?
 ユーザ2: what is **the baseball player** doing?
 システム: he is throwing a ball.

- 画像キャプション生成 [Vinyals+ 2015]



出力: a **costumed girl** stands near a shelf.
 正解: a **girl in a cat costume** is standing near shelving.
 評価値 = 0.95

2. 視覚に基づく言い換え (VGP)

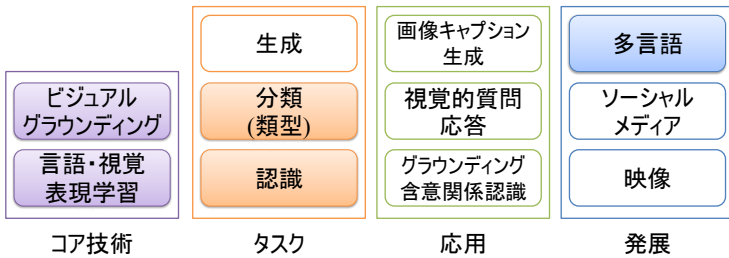


Caption 1: a baseball player in a red jersey throwing a ball at the pitchers mound .
 Caption 2: a baseball team pitcher throwing a ball to the batter .
 Caption 3: a little league pitcher in a red shirt .
 Caption 4: a male is standing on a base pitching a ball .
 Caption 5: the pitcher is wearing a red uniform shirt .

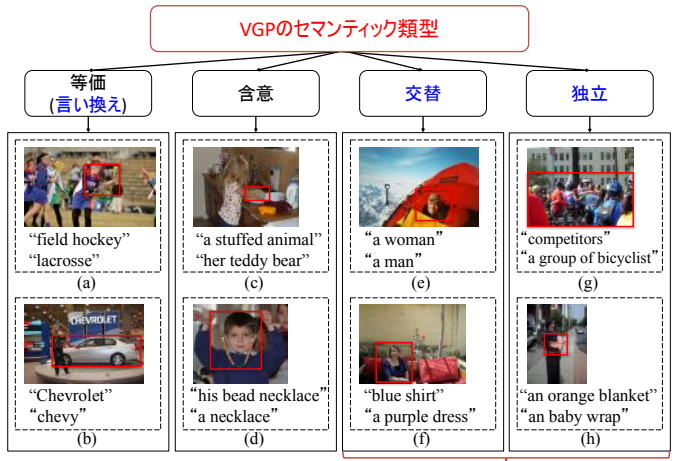
VGP set 1 a baseball player a baseball team pitcher a little league pitcher a male the pitcher	VGP set 2 a red jersey a red shirt a red uniform shirt VGP set 3 the pitchers mound	VGP set 4 a base VGP set 5 a ball
---	--	--

VGP: Visually Grounded Paraphrase

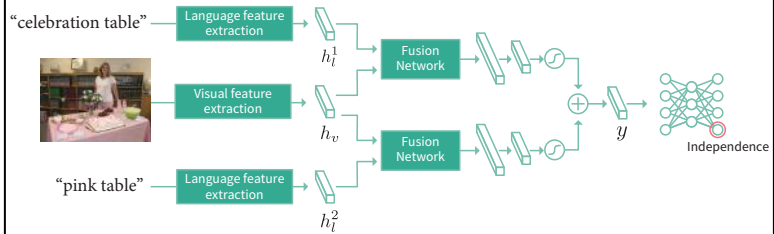
3. VGP研究の全体像



4. VGPのセマンティック類型と分類モデル

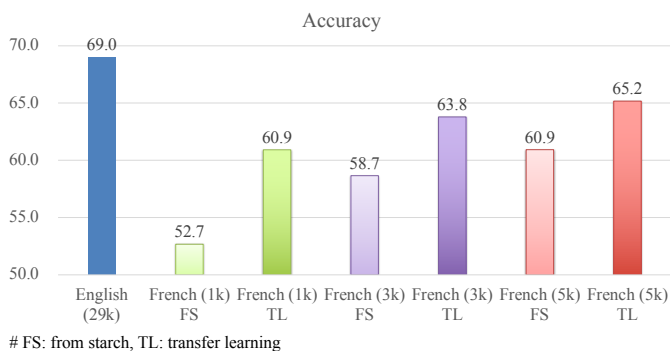
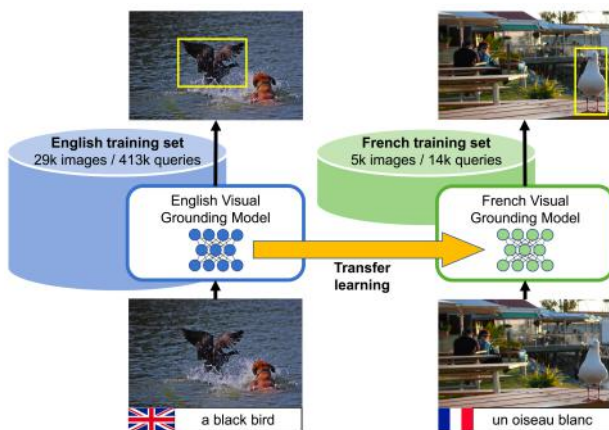


新しいセマンティック現象の発見



VGPを分類するには言語・画像表現の同時学習が重要である

5. 言語横断ビジュアルグラウンディング



6. まとめと今後の予定

まとめ

- VGPという新しい言語・視覚の概念の提案
- VGPという研究分野を整理
- VGP類型と言語横断ビジュアルグラウンディングの研究

今後の予定

- VGP生成
- VGPを視覚的質問応答や画像キャプション生成へ応用
- VGP研究をソーシャルメディアや映像データへ発展
- 最終的にはVGPという研究分野を確立