画像をピボットとしたパラフレーズの抽出による自然言語と画像理解の高度化
（言い換え認識による言語と視覚理解）

大阪大学　データビリティフロンティア機構
特任助教　チョ　シンキ (CHU, Chenhui)

1.言語と視覚理解における言い換え認識の重要性

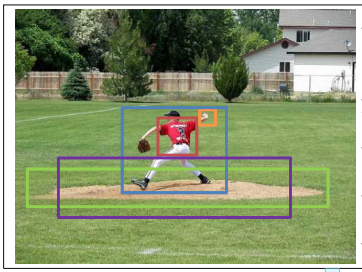
- 視覚的質問応答 [Wu+ 2017]

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) [COLING 2018]



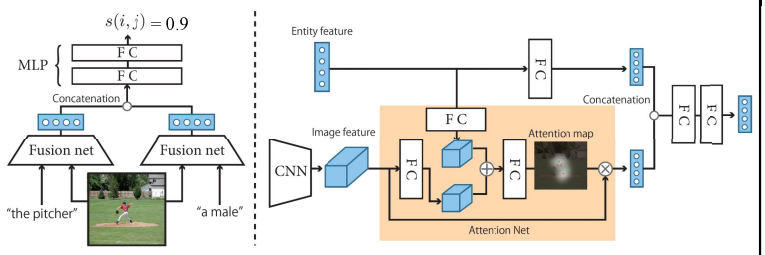
Caption 1: **a baseball player** in **a red jersey** throwing **a ball** at **the pitchers mound** .
Caption 2: **a baseball team pitcher** throwing **a ball** to the batter .
Caption 3: **a little league pitcher** in **a red shirt** .
Caption 4: **a male** is standing on **a base** pitching **a ball**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

⇩

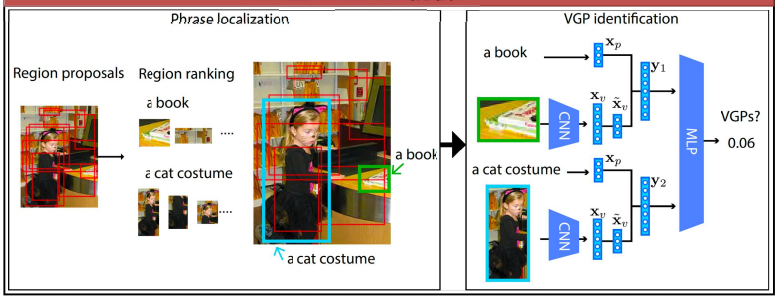
VGP set 1	VGP set 2	VGP set 4
a baseball player	a red jersey	a base
a baseball team pitcher	a red shirt	
a little league pitcher	a red uniform shirt	VGP set 5
a male	VGP set 3	a ball
the pitcher	the pitchers mound	

VGP: Visually Grounded Paraphrase

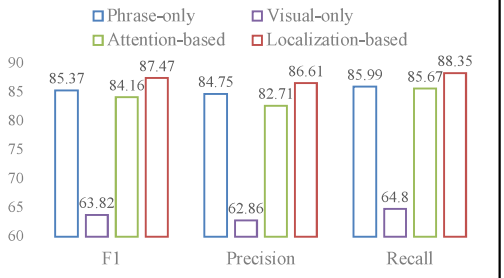
3. Attentionに基づくVGP認識 [COLING 2018]



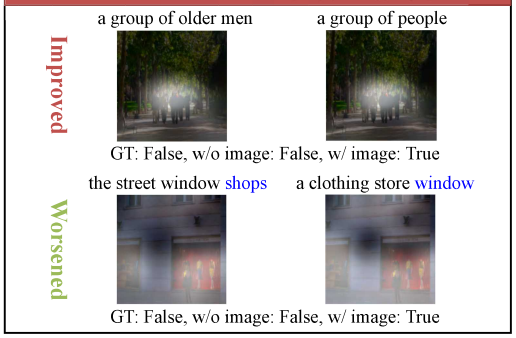
4. Phrase Localizationに基づくVGP認識 [CVPR 2018 Workshop]



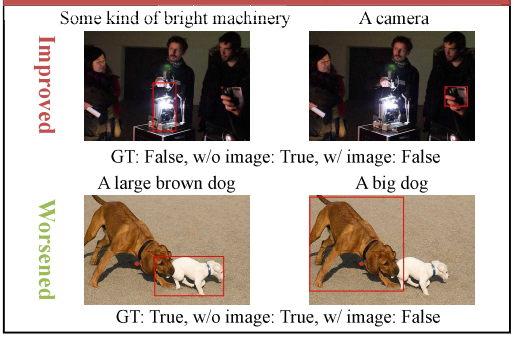
5. VGPの認識の結果



6. Attentionに基づくVGP認識の結果例



7. Localizationに基づくVGP認識の結果例



8. VGPのセマンティック類型

9. まとめと今後の予定

まとめ

- VGPのという新しい言語と視覚の概念の提案
- Attentionに基づくVGP認識モデルの提案
- Phrase localizationに基づくVGP認識モデルの提案

今後の予定

- VGP認識モデルの改善
- 視覚的関係を含む任意のVGPの認識モデル
- VGP現象の解明とセマンティック類型の構築
- SNSなどの実データでのVGP認識

10. 発表文献

[1] Chenhui Chu, Mayu Otani and Yuta Nakashima. iParaphrasing: Extracting Visually Grounded Paraphrases via an Image. In Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018), pp.3479–3492, (2018).

[2] Mayu Otani, Chenhui Chu, Yuta Nakashima. Visually Grounded Paraphrase Extraction via Phrase Grounding. Workshop on Language and Vision at CVPR 2018, (2018).