

MMC-GAN [Aly+, RA-L18]

✂ carry-and-placeタスクに焦点

✂ 曖昧な自然言語による指示に対するtarget areaの可能性を予測

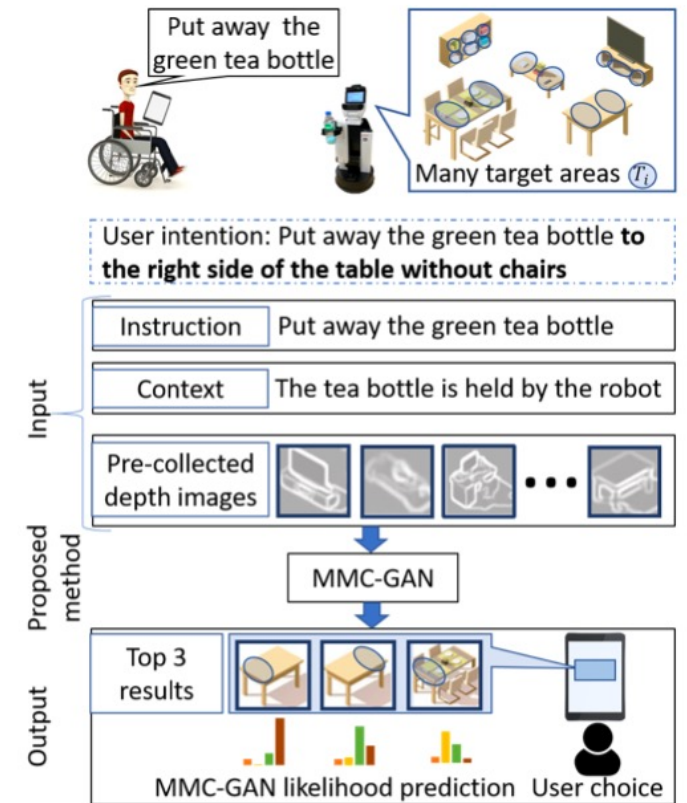
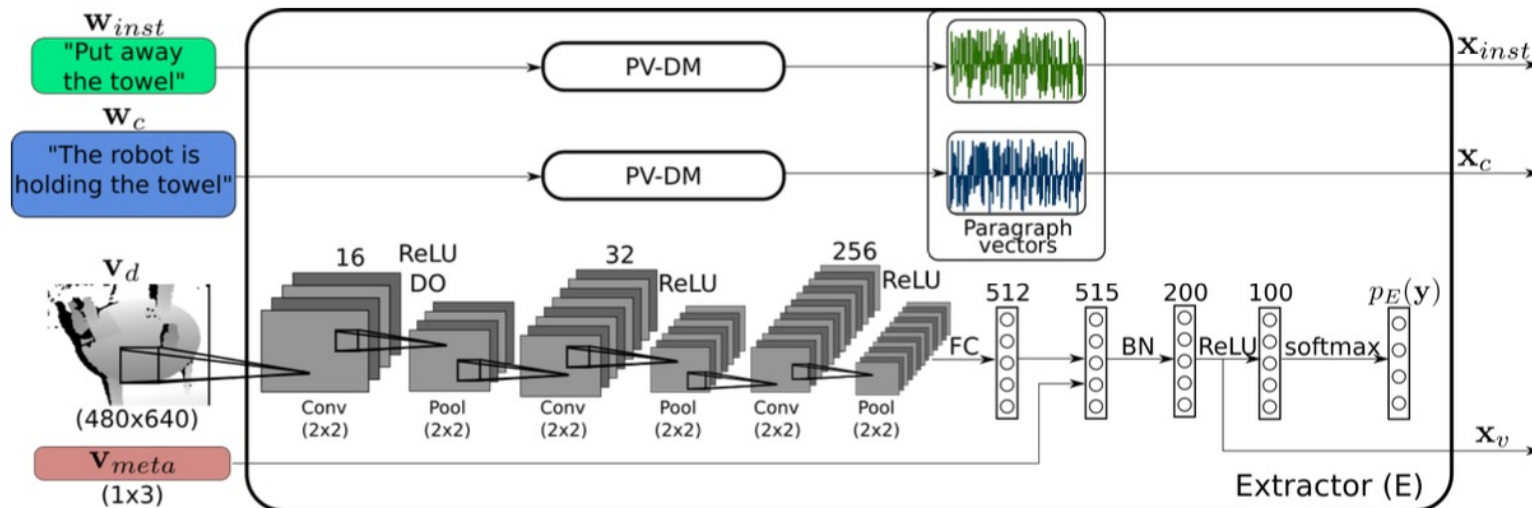
✂ GANによるデータ拡張

✂ 環境やロボットの配置等が多様

✂ 大規模なデータセットが収集不可

✂ マルチモーダルな入力に対応

✂ 言語をPV-DM、画像をCNNで特徴抽出



Aly Magassouba, Komei Sugiura and Hisashi Kawai. (2018). A Multimodal Classifier Generative Adversarial Network for Carry and Place Tasks from Ambiguous Language Instructions. arXiv:1806.03847

MTCM-AB [Aly+, RA-L19]

✂ MTCM(multimodal target-source classifier model)を
ABN(attention branch network)を用いて拡張

✂ MLU-FI(multimodal language understanding for fetching instructions)に焦点

✂ 曖昧な自然言語による指示からユーザの意図を正確に推論

✂ マルチモーダルなattention map

✂ TAB

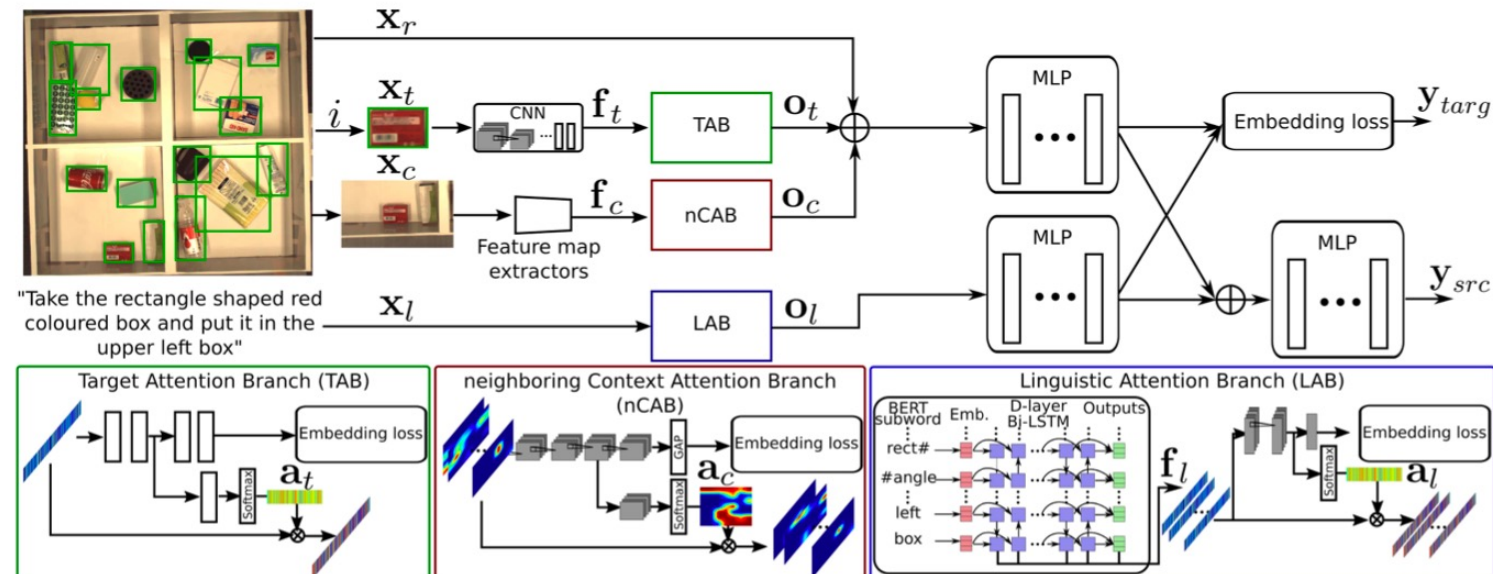
✂ ターゲット画像の候補に対する
Attention Branch(AB)

✂ nCAB

✂ ターゲットの周囲を含む
画像に対するAB

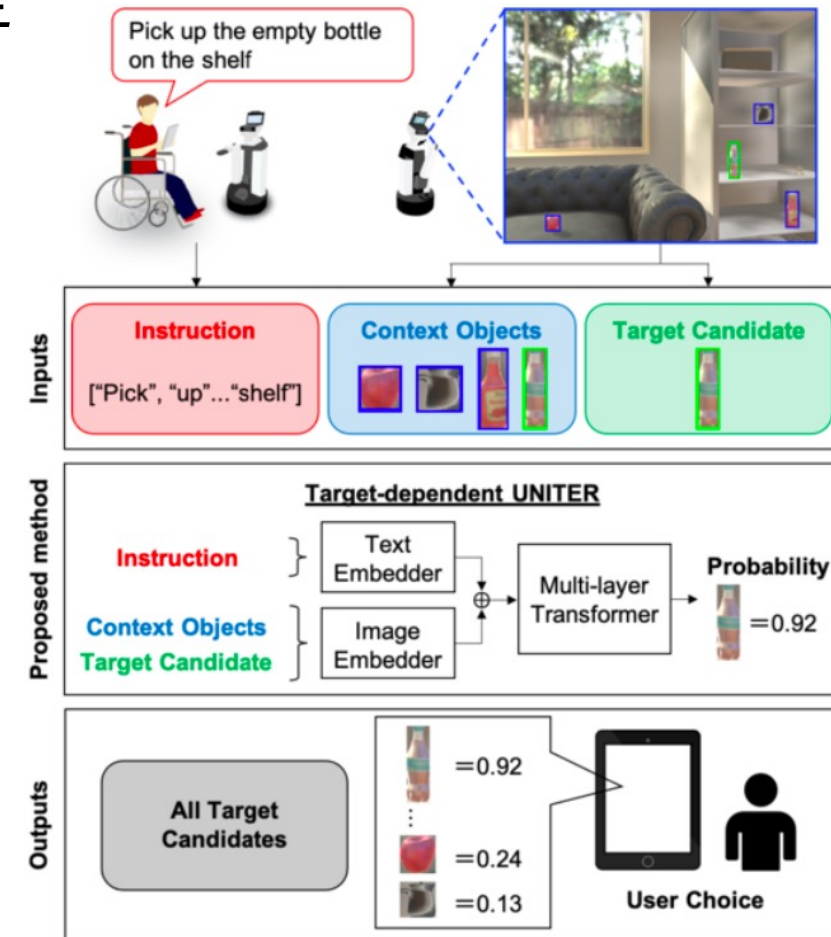
✂ LAB

✂ 言語入力に対するAB



Target-dependent UNITER[Ishikawa+, IROS21]

- ✖ 曖昧性を含む自然言語による命令文から対象物体を特定
- ✖ 画像とテキストの共同理解にUNITERを採用
 - ✖ target候補を入力として新たに導入
- ✖ Image Embedder、Text Embedder、Transformerから構成
 - ✖ Image、Textの特徴量を連結し入力
 - ✖ 画像中の各物体とテキスト中の各単語の関係性を学習



Shintaro Ishikawa and Komei Sugiura. (2021).
Target-dependent UNITER: A Transformer-Based Multimodal
Language Comprehension Model for Domestic Service
Robots.