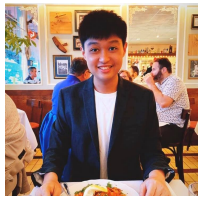
SSCR: Iterative Language-Based Image Editing via Self-Supervised Counterfactual Reasoning



Tsu-Jui Fu

Xin Wang

Scott Grafton

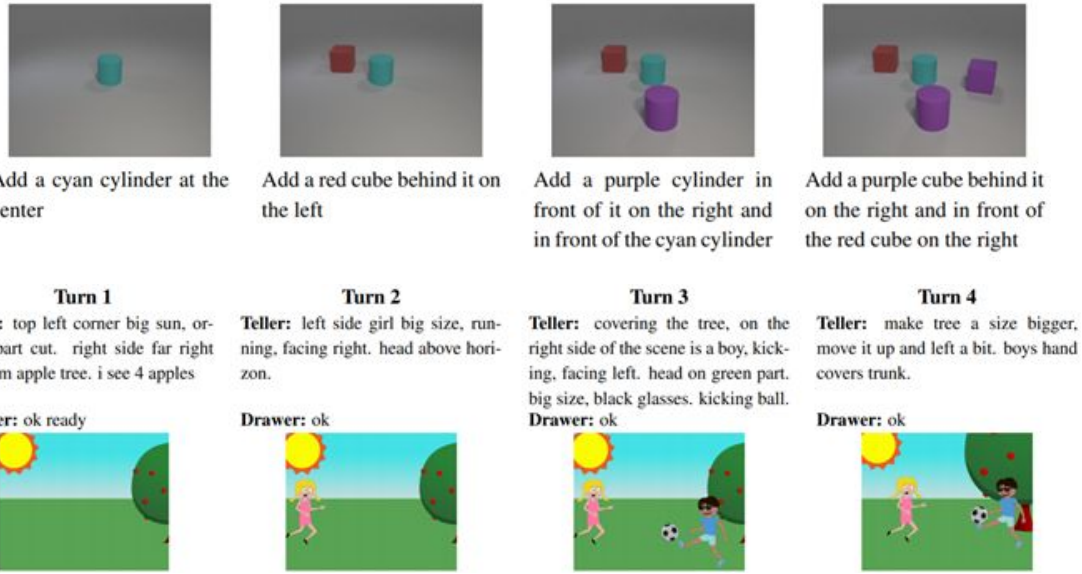
Miguel Eckstein

William Wang

UC Santa Barbara

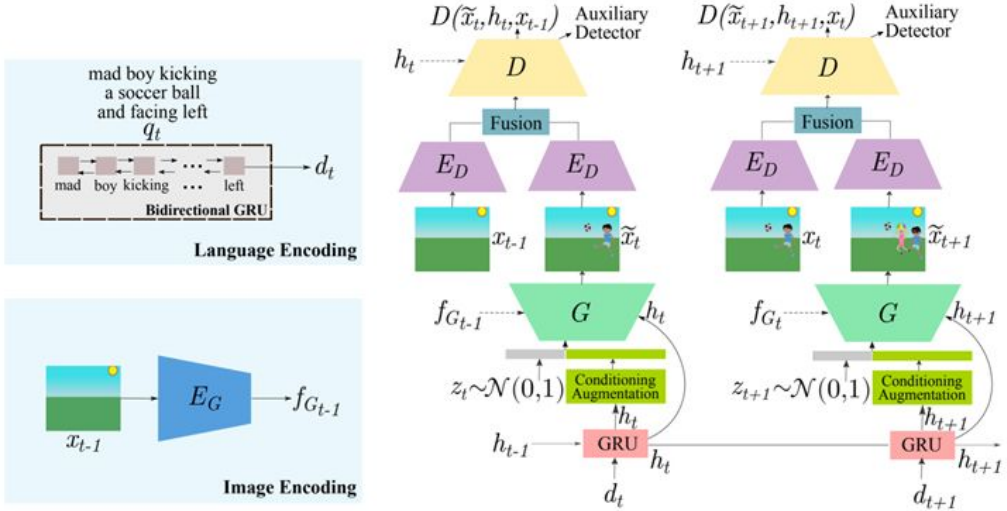
LBIE

- **Language**-based image editing (LBIE)
  - **Iterative** (step-by-step) editing



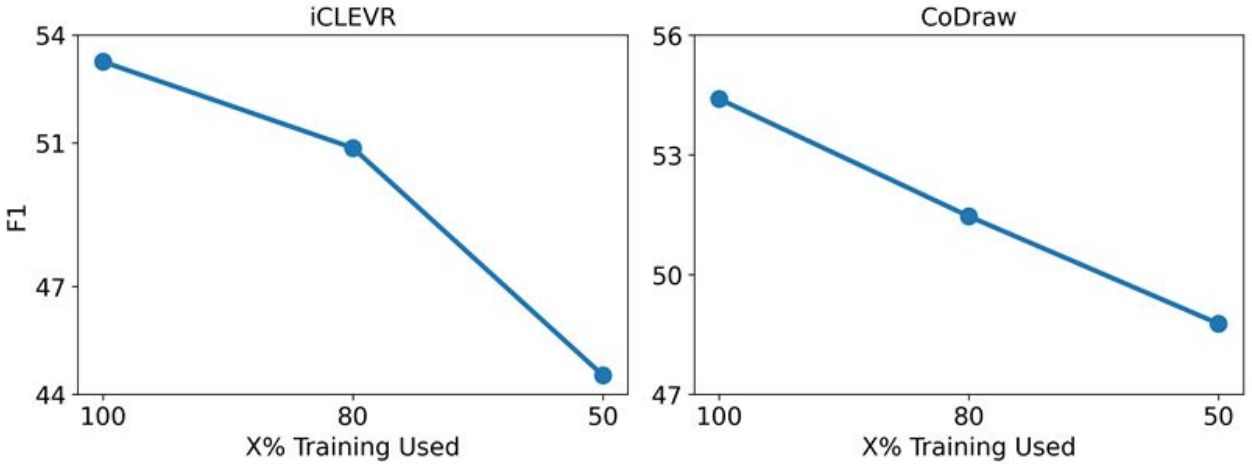
GeNeVA (Baseline)

- Conditional GAN
  - G conditioned on h_t and f_{t-1}
  - D as binary classifier and conditioned on h_t



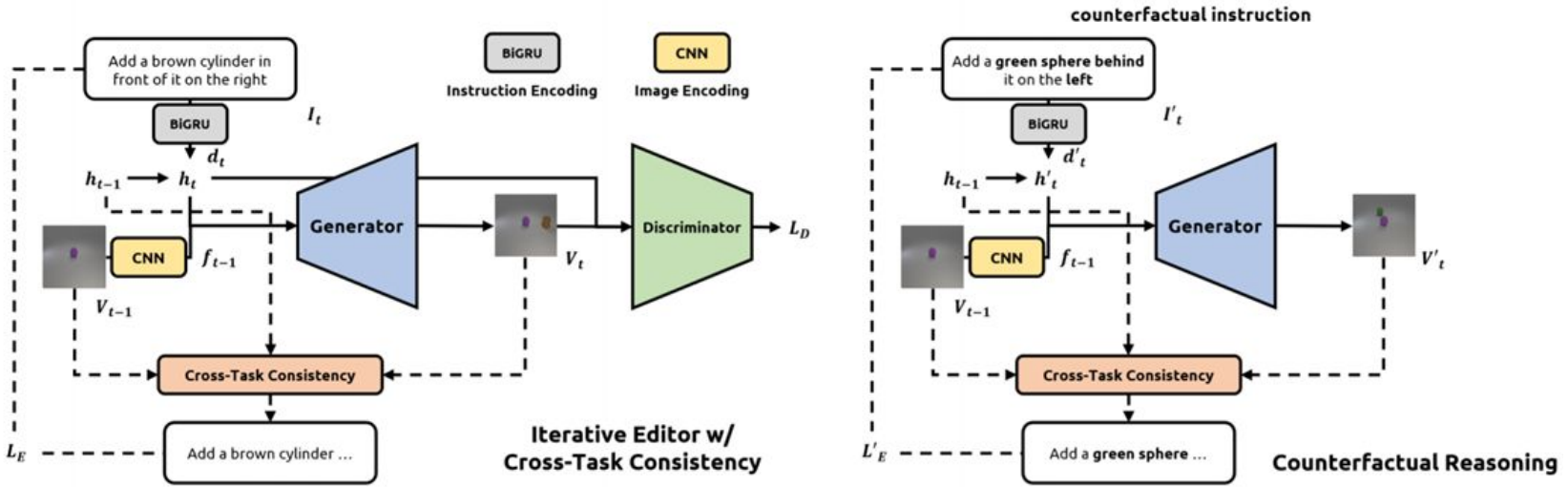
GeNeVA (Baseline)

- Suffers from **data scarcity**
  - D can only provide **binary** signal but **not explicit**
  - **Not enough data** for D to train G



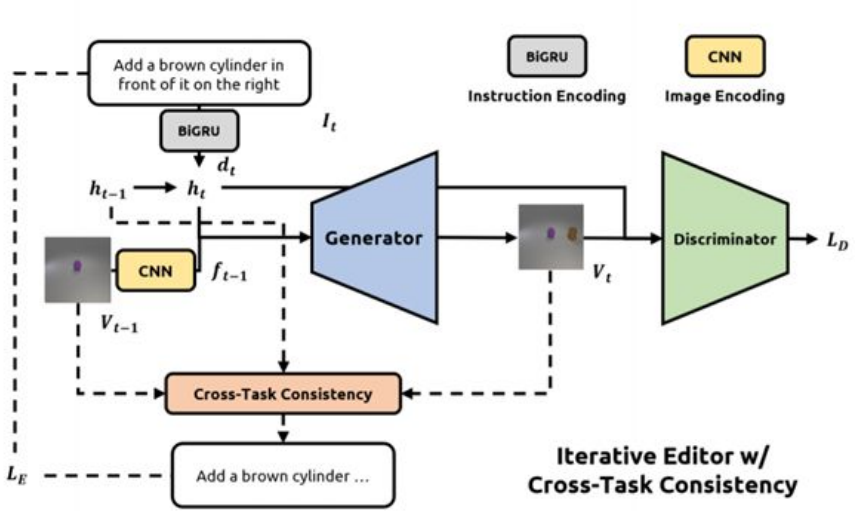
SSCR (Ours)

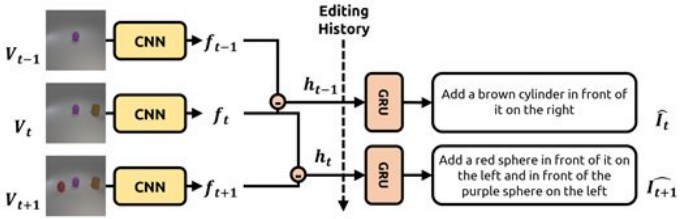
- Self-Supervised counterfactual reasoning (SSCR)
  - Cross-Task consistency (CTC) to provide **token-level loss**
  - Considers various **counterfactual instructions**



SSCR (Ours)

- Cross-Task consistency (CTC)
  - ***CTC***: describes the difference iteratively
  - L_E: token-level loss





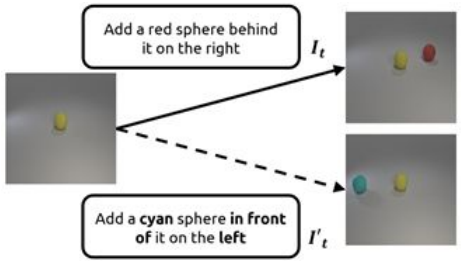
$$g_0 = [f_d, h_{t-1}],$$
$$\hat{w}_i, g_i = \text{GRU}(w_{i-1}, g_{i-1}),$$
$$\hat{I}_t = \{\hat{w}_1, \hat{w}_2, ..., \hat{w}_L\},$$
$$L_E = \sum_{i=1}^{L} \text{CELoss}(\hat{w}_i, w_i)$$

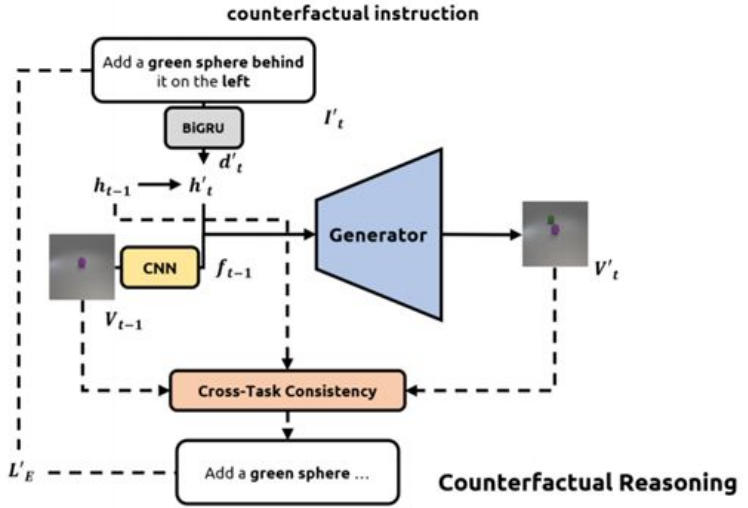
SSCR (Ours)

- Counterfactual reasoning
  - Does **intervention** for counterfactual instruction ***I'***
  - Applies CTC to **train self-supervisedly** by L'_E

Dataset	Token Type	Example
i-CLEVR	color object relation	blue, purple cylinder, cube at the center, in front of
CoDraw	size object relation	small, meidum sun, boy in the middle, on the left

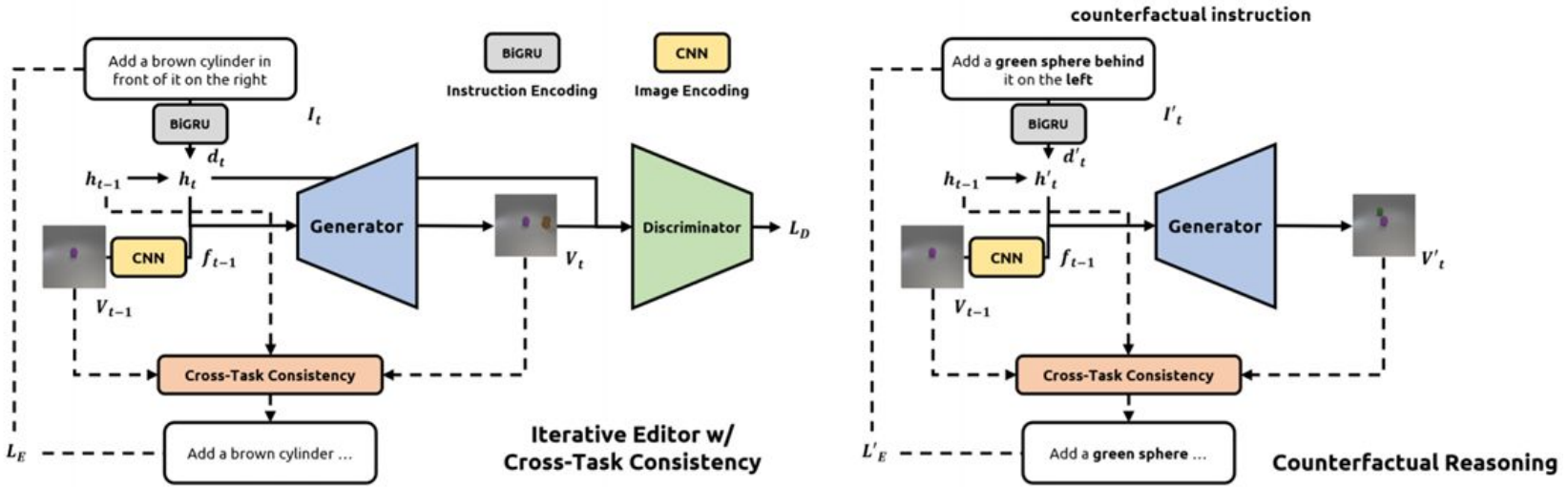
replace ***token*** to do intervention





SSCR (Ours)

- Cross-Task consistency (CTC)
  - Provides explicit **token-level loss** (L_E)
  - Trains counterfactual instructions **self-supervisedly** (L'_E)

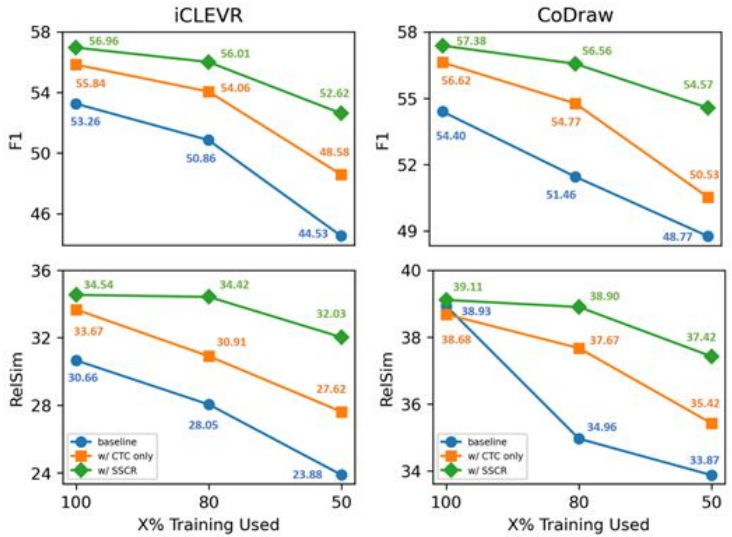


Experiments

- Evaluation Metrics
  - Precision / Recall / **F1**
    - Matches **objects** between prediction and groundtruth
  - **RelSim**
    - Considers both **objects** and **related position**

Experiments

- Ours vs Baseline under **data scarcity**
  - Baseline drops a lot
  - SSCR achieves **similar performance** when using **only 50%**

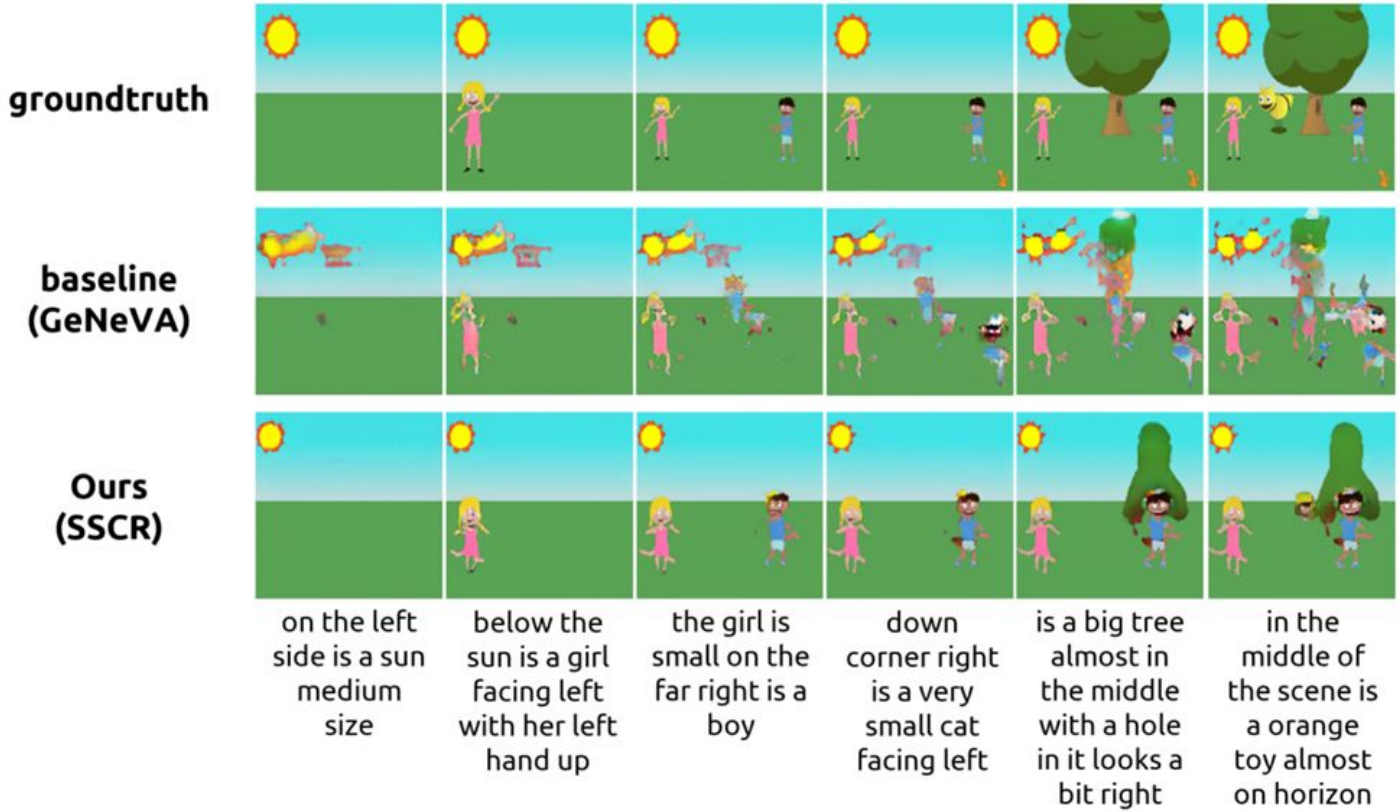


PPL & **BLEU** for CTC

X% Training Used	PPL	BLEU
100%	0.1073	50.236
80%	0.1295	48.873
50%	0.1163	48.763

CTC provides **meaningful training loss** for SSCR even under data scarcity

Visualization Examples



https://tsujuifu.github.io/projs/emnlp20_sscr.html