**1. Introduction**

-

**2. Related Work**

- Scene Graph Generation model은 크게 two-stage model과 one-stage model로 분류할 수 있음

- two-stage model은 object detection model과 relation prediction model을 순차적으로 따로 학습하며, object detector에서 출력한 N개 object 간의 모든 object pair를 relation prediction model에 입력해 relation을 예측하는 방식임

- 높은 relation extraction 성능을 보이지만, 연산량이 많아지며 부정확한 object detection이 다음 과정까지 전파되는 문제가 있음

- 이를 개선하기 위해 제안된 one-stage model은 세 가지로 분류할 수 있으며, 최근에는 대부분 DETR과 같은 one-stage object detection model을 활용함

- Object-triplet detection model은 object detector 위에, subject-predicate-object로 구성된 triplet query로부터 relation을 예측하는 triplet predictor를 추가한 구조임

- 이 경우 triplet predictor가 object detector로부터 얻는 사전 정보가 없기 때문에, detected object로부터 직접 정보를 추출하는 module을 포함하다 보니 복잡한 구조를 갖게 됨

- Triplet detection model은 object detector 없이 triplet detector만으로 이루어져 있음

- triplet을 직접 찾기 때문에, relation을 갖지 않는 object는 놓쳐 complete scene graph가 아닌 sub-graph를 감지하는 것에 집중하게 되는 단점이 있음

- Relation extraction model은 object detector와 triple query를 사용하지 않는 작은 relation predictor로 이루어져 있으며, object detector에서 relation 관련 정보를 추출해 사용함

- Relationformer는 transformer decoder에 relation token을 추가해 global information을 추출한 후, object query pair와 relation token의 final hidden representation을 결합해 relation predictor에 입력함

**3. Method**

-