

# Bounding Boxを付与したフローグラフコーパスの提案

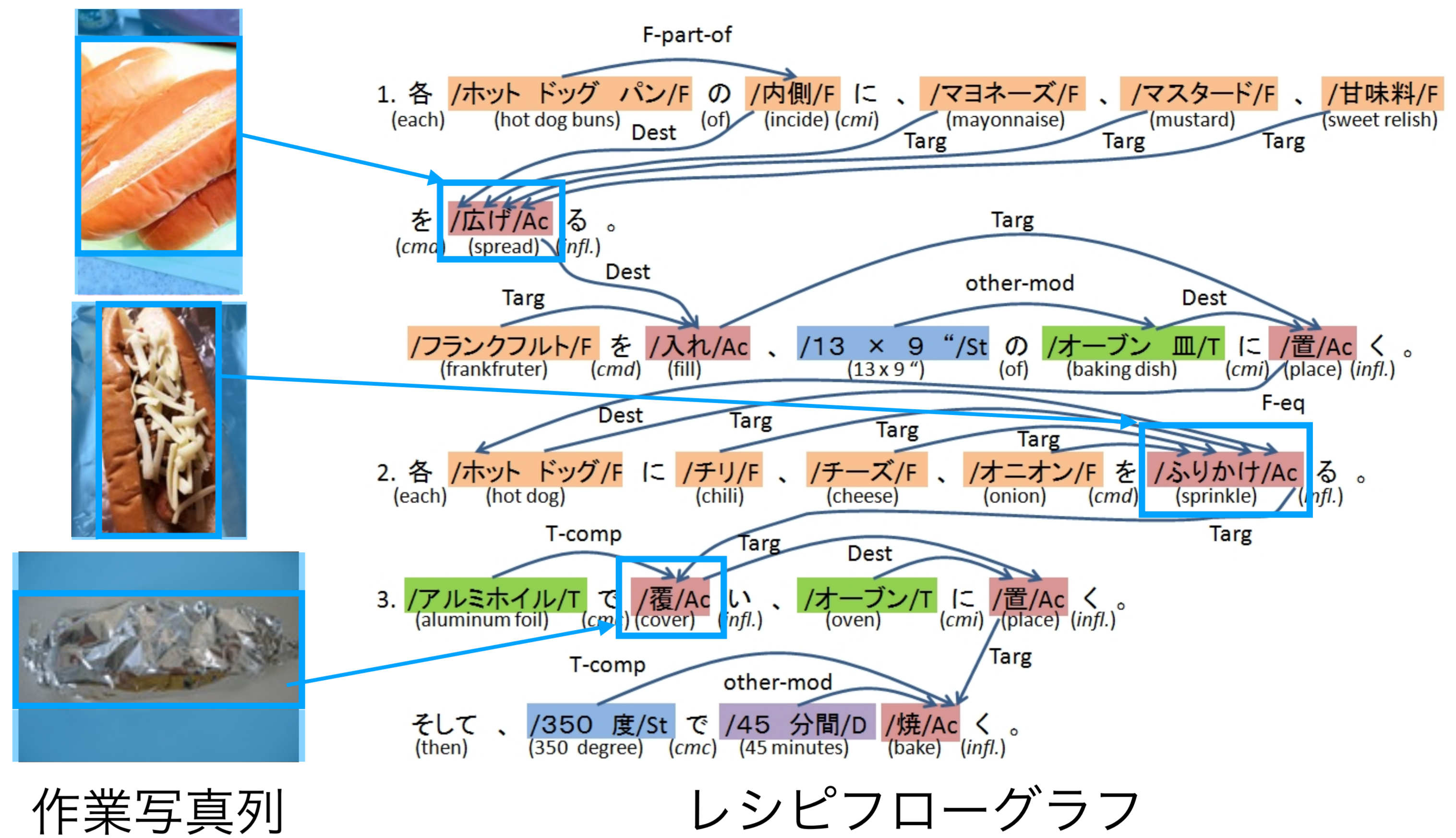
西村 太一<sup>1</sup> 橋本 敦史<sup>2</sup> 原島 純<sup>3</sup> 山肩 洋子<sup>4</sup> 森 信介<sup>1</sup>

京都大学<sup>1</sup> オムロンサイニクエクス株式会社<sup>2</sup> クックパッド株式会社<sup>3</sup> 東京大学<sup>4</sup>

## 研究概要

- 手順書においては、手順を経て物体が視覚的に変化
- 直接言語化が難しい中間状態

フローグラフコーパス [Mori+ LREC14]  
各レシピ固有表現へ BoundingBox を付与



## アノテーション結果

### ①アノテーション方法

- フローグラフコーパスから全手順に写真が付与しているものを抽出
- 70レシピについて1名がアノテーション
- F(食べ物), T(道具), Ac(動作)の3つにNEに付与可能
- Ac(動作)には, “動作中”, “動作済”, “判別不能”の3クラスを付与

step 1

アノテーション画面

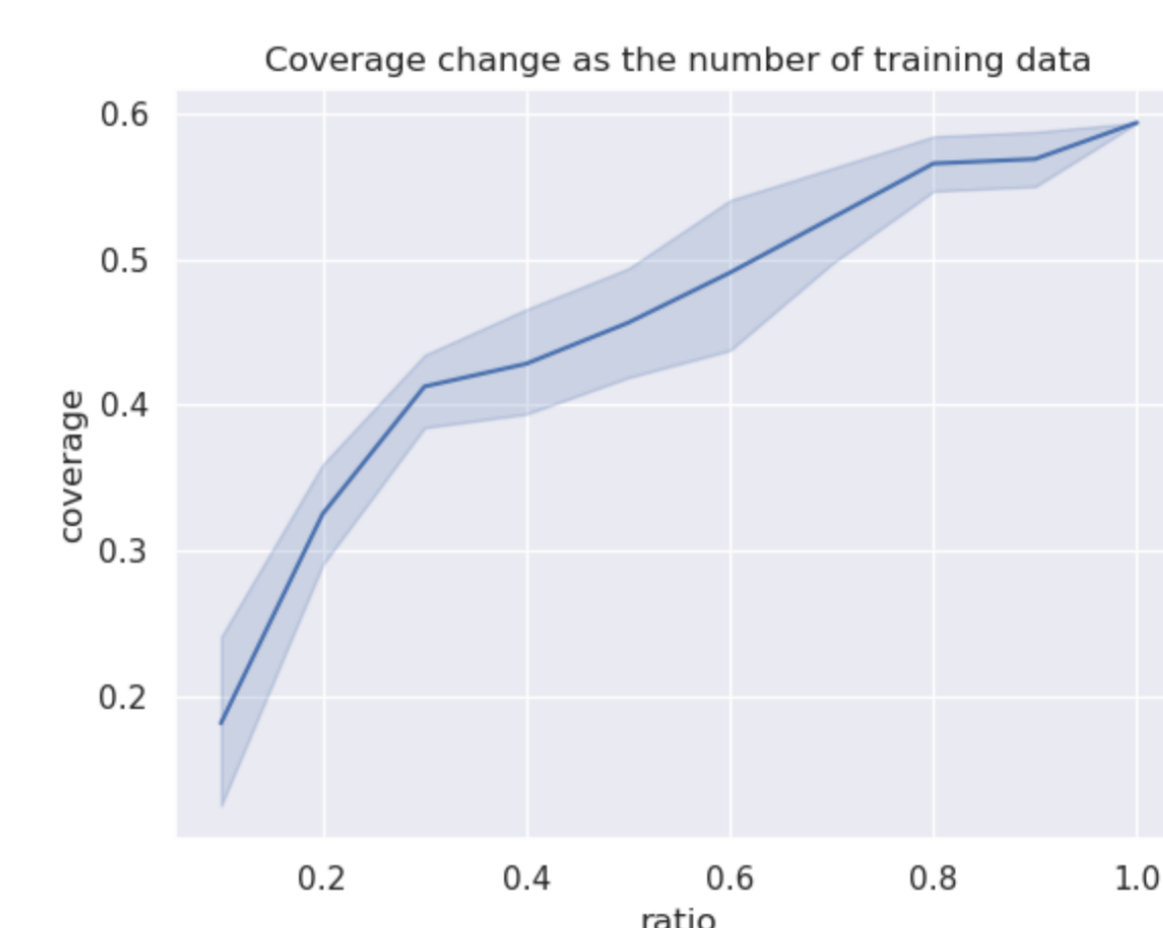
### ②アノテーション後の統計情報

	手順	文	単語	NE	F	T	Ac	BB
70 レシピ	3.9	7.1	96.5	39.9	13.3	4.2	12.7	4.4

1レシピあたりの統計情報 ※BB=Bounding Box

### ③NEのカバレッジ

- train/testにランダムに8:2に分割
- testの各NEがtrain中に存在するかを計算



上昇傾向  
よりデータを増やすことで  
カバレッジは上げられる

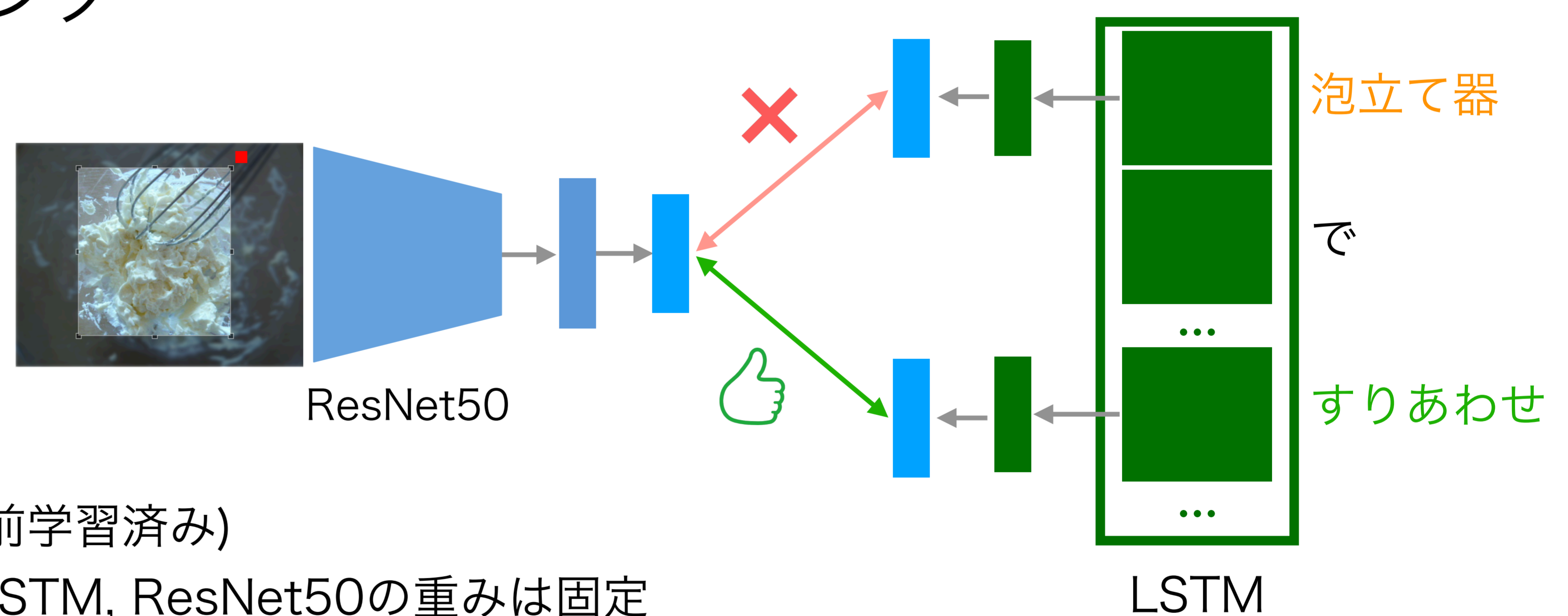
## 予備実験 | シンボルグラウンディング

### ①問題設定

- 各 Bounding Box を入力とし, 手順中のどの NE に付与しているか当てる

### ②実装

- Triplet margin loss [Balntas+ BMVC16]
- margin : 0.1
- LSTMの次元数512, ResNet50の出力2048 (事前学習済み)
- それぞれ512次元になるように全結合層追加, LSTM, ResNet50の重みは固定



### ③実験結果

- Recall@k, medR
- 学習データ数を変化させた時の Recall@1 の推移

	medR	Recall@1	Recall@2	Recall@3
ランダム	4	0.12	0.24	0.36
フロー グラフなし	1	0.51	0.63	0.66



データ数とともに性能は上昇傾向

## 応用と今後の課題

- 物体の状態を含めた Object Tracking
  - フローグラフも同時に使うことで, 物体の構成要素も考慮
- キャプション生成
  - アノテーションされた矩形と NE からより画像に合ったキャプション生成
- マルチモーダル手順書理解, Entity State Tracking
  - 各 Entity の状態追跡 [Bosselut+ ICLR2018], フローグラフ生成 [Maeta+ IWPT2015]