Accepted to CVPR 2020

# Sketch-BERT

## : Learning Sketch Bidirectional Encoder Representation from Transformers by Self-supervised Learning of Sketch Gestalt

손수현

- Sketch image를 활용한 task

1. Recognition

   : sketch 그림의 category(label)을 예측하는

2. Retrieval

   : query sketch가 주어졌을 때, 그와 같은 category의 sketch를 예측하는

3. Sketch Gestalt *본 논문에서 처음으로 제안한 task

   : masked된 부분을 예측하여 realistic sketch image를 생성하는

→ Sketch is stored as a sequence of data points, a vector format representation

기존의 모델(SketchRNN)이 학습하는 representation은 generation task에서만 중요했고,
다른 task에서는 크게 도움을 주지 못함

↓

SketchBERT

: NLP의 BERT 구조에서 embedding과 pretraining algorithm에 변화를 주어 다양한 sketch task를 수행하도록

→ Sequential data를 modeling할 때 적합한 구조

sketch can be represented as a list of points, where each point contains 5 attributes

$$(\Delta x, \Delta y, p_1, p_2, p_3)$$

위치정보 $\Delta x, \Delta y$ : relative offsets between current point and previous point

$p_1$ = 1 : represents the other sequential points of sketches

상태정보 $p_2$ = 1 : indicates the ending of one stroke

$p_3$ = 1 : means the ending of the whole sketch

sketch can be represented as a list of points, where each point contains 5 attributes

$$(\Delta x, \Delta y, p_1, p_2, p_3)$$

$\Delta x, \Delta y$ : relative offsets between current point and previous point

$p_1$ = 1 : represents the other sequential points of sketches

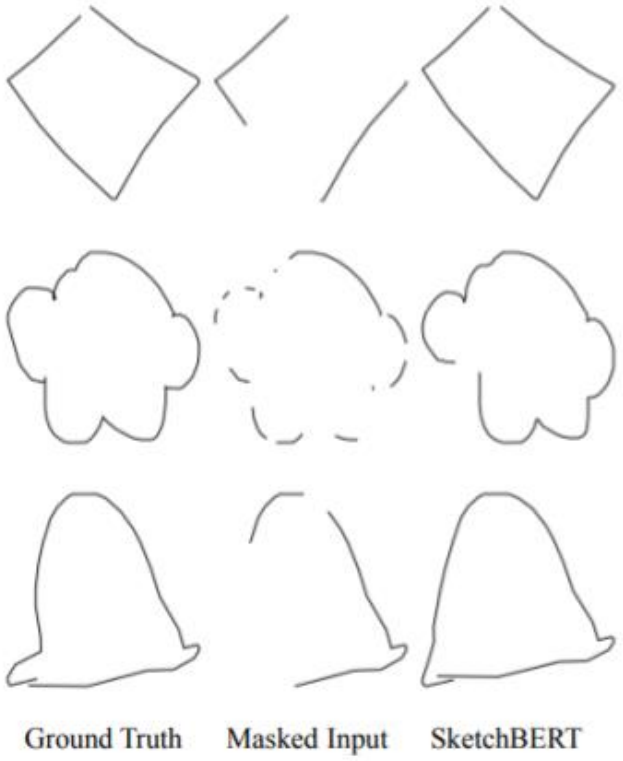
$p_2$ = 1 : indicates the ending of one stroke

$p_3$ = 1 : means the ending of the whole sketch

Point Embedding
: 5 attributes에 대한
+
Positional Embedding
:각 sequential point 의 position에 대한
+
Stroke Embedding
: 한 stroke(획)의 sequence에 대한

→ BERT의 input으로

# Self-Supervised Learning by Sketch Gestalt




- **Sketch Gestalt (sGesta)**

  **: 본 논문에서 제안한 새로운 self-supervised learning process**

  : recovering the masked points in sketches

  ↳ 이를 수행하는 SGM(Sketch Gestalt Model) 제안

- **SGM**

  : 5개의 key information으로 표현된 point는 SGM에 의해 mask되고 predict된다.

  : 2가지의 mask strategies 제안

  (15%)

  1) positional offset

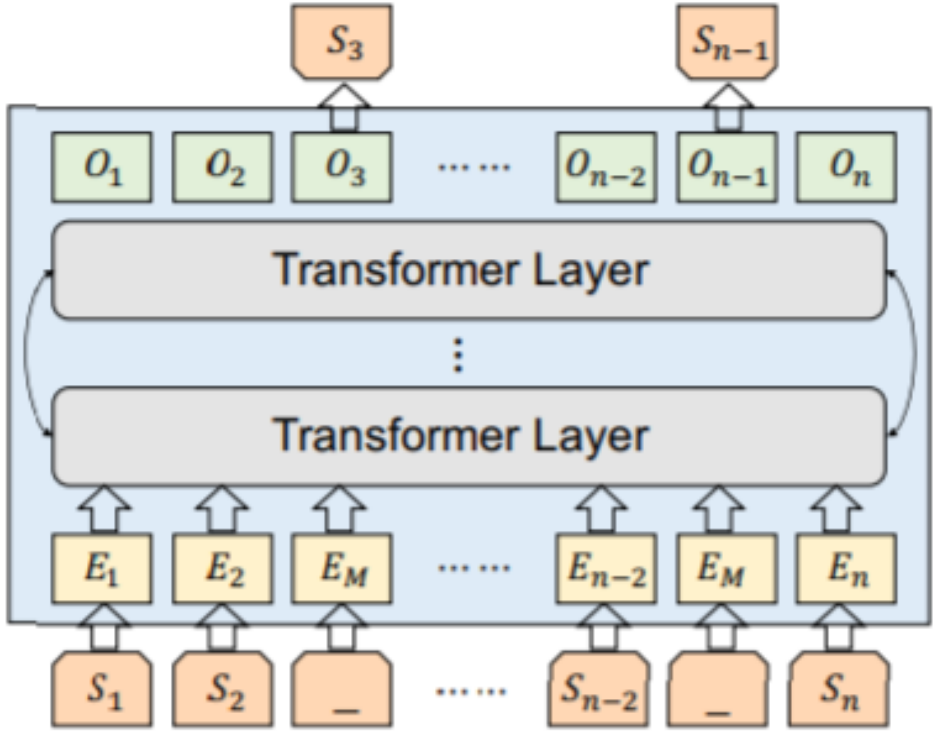
  2) state information

$$(\Delta x, \Delta y \; p_1, p_2, p_3)$$

## Self-Supervised Learning by Sketch Gestalt

1. Mask Position Prediction
   - Δx, Δy 값을 0으로 함으로써 masking

2. Mask State Prediction
   - p1,p2,p3값을  0을로 함으로써 masking




(a) Sketch-BERT for Sketch Gestalt

**downstream tasks after the pre-training procedure**

Sketch Recognition
Sketch Retrieval
Sketch Gestalt

- 2가지 Dataset을 사용하여 training & test

: QuickDraw dataset

(50 million, 345 classes of common object)

: TU-Berlin dataset

(보다 quality가 좋은, 250개의 category, 각 category 별 80개의 sketches)

| Models | QuickDraw | | | TU-Berlin | | |
|---|---|---|---|---|---|---|
| | Top-1 (%) | Top-5 (%) | mAP (%) | Top-1 (%) | Top-5 (%) | mAP(%) |
| Bi-LSTM [9] | 70.91 | 89.52 | 60.11 | 31.40 | 59.60 | 23.71 |
| Sketch-a-Net [27] | 74.88 | 90.10 | 65.13 | 37.25 | 63.50 | 26.18 |
| DSSA [22] | 78.16 | 91.04 | 68.10 | 38.45 | 66.10 | 28.77 |
| ResNet18 [8] | 80.34 | 91.71 | 70.98 | 41.45 | 67.10 | 29.33 |
| ResNet50 [8] | 82.41 | 92.52 | 74.84 | 51.80 | 74.45 | 36.94 |
| TCNet [20] | 83.59 | 92.57 | 76.38 | 55.30 | 79.45 | 38.78 |
| Sketch-BERT (w./o.) | 63.13 | 84.70 | 55.10 | 32.50 | 57.90 | 24.14 |
| Sketch-BERT (w.) | **85.47** | **93.49** | **78.87** | **57.25** | **81.50** | **41.54** |

sketch gestalt 포함 유무