Supplementary Material - Dynamic Graph Generation Network: Generating Relational Knowledge from Diagrams

Daesik Kim[1,2] YoungJoon Yoo[3] Jeesoo Kim[1] Sangkuk Lee[1,2] Nojun Kwak[1]
[1]Seoul National University [2]V.DO Inc. [3]Clova AI Research, NAVER Corp.
{daesik.kim|kimjiss0305|sangkuklee|nojunk}@snu.ac.kr youngjoon.yoo@navercorp.com

1. Results of Recall@K

While we measured AP for evaluation in this paper, we additionally utilized recall metric for measuring retrieval power of relationships due to the sparsity of the relationship. Table 1 shows results of Recall@k metric (R@k) on AI2D test dataset. The R@k measures the fraction of ground-truth relationship that appears among the top-k most confident predictions. The results of R@5, R@10 and R@20 demonstrate similar trend of results compared to those of mAPs.

Table 1. Comparison results of Recall@K on the AI2D test set.

Method	R@5	R@10	R@20
Vanilla GRU	21.79	33.97	48.87
DGGN			
w/o global feature	21.44	33.63	49.18
w/o weighted mean pool	22.62	35.75	51.60
w/ ROI-pooled feature	21.45	34.21	49.87
DGGN	**22.66**	**35.93**	**51.73**

2. Training Details

For training, We jointly optimized the overall loss of the proposed algorithm with ADAM optimizer with default parameters ($\beta_2 = 0.999, \epsilon = 10^{-9}$). For the three losses in overall loss (1), we set $\alpha = 0.2$, $\beta = 0.1$ and $\gamma = 1.0$. The initial learning rate is set to $1 \times \epsilon^{-4}$ and is multiplied by 0.09 in every 1000 iterations. The batch size is set to 32 and we evaluated our model after 15000 iteration (≈150 epochs).

$$L = \alpha L_c + \beta L_l + \gamma L_r. \tag{1}$$

3. Details of Post-processing

In this section, we explain a detailed post-processing procedure of the proposed method. Once relationships are determined among objects, we can additionally make new relationship between objects sharing the same intermediate node. For example, given two text objects sharing the same blob object, we can say that one text is linked to another one. Also, given two text objects connected by two intermediate blob objects, we can say equivalently to the previous case. In most case, the text object represents the name or explanation of the connected blob object. Consequently, making further connections by this rule-based algorithm, we can generate sentences using given texts. Using extensively connected texts, we just put an additional phrase of "links to" such as "Lavar links to Fly". Note that localized text boxes are recognized using Tesseract[1]. Algorithm 1 shows details of post processing.

[1]https://github.com/tesseract-ocr/tesseract

Algorithm 1 Post processing algorithm

Require: Relation set R generated by the proposed DGGN
Ensure: Generated sentences set S

$S \leftarrow \emptyset$
repeat
  $R_a \leftarrow \{o_{a1}, o_{a2}\} \in R$
  $R_b \leftarrow \{o_{b1}, o_{b2}\} \in R$
  if $R_a \cap R_b \in$ '$text$' **then**
    Continue
  else if $R_a \cap R_b \in$ '$blob$' **then**
    if $R_a - R_b \in$ '$text$' & $R_b - R_a \in$ '$text$' **then**
      Generate sentence S_{ab}
      $S \leftarrow S \cup S_{ab}$
    else if $R_a - R_b \in$ '$text$' & $R_b - R_a \in$ '$blob$' **then**
      Find R_c satisfying $\{R_c \cap R_b \in$ '$blob$' & $R_c - R_b \in$ '$text$' $\}$
      Generate sentence S_{ac}
      $S \leftarrow S \cup S_{ac}$
    end if
  end if
until all elements in R are visited

After generating sentences about relationship, we added sentences about facts of detected elements in a diagram. As shown in Figure 4, counts of objects and stages, and names of elements are added to help richer descriptions about a diagram.

4. Additional Qualitative Results

In next pages, we present additional qualitative results on diagram graph generation and question answering. We provide results of diagram graph generation of various layouts and topics as depicted from Figure 1 to Figure 3. The results of DGGN are compared with those of those of vanilla GRU. Ground truths are also shown. In Figure 4, we also show pipelines from diagram graph to question answering with post-processing described in the previous section. For two different types of questions, *relationship* and *count*, related sentences are highlighted.

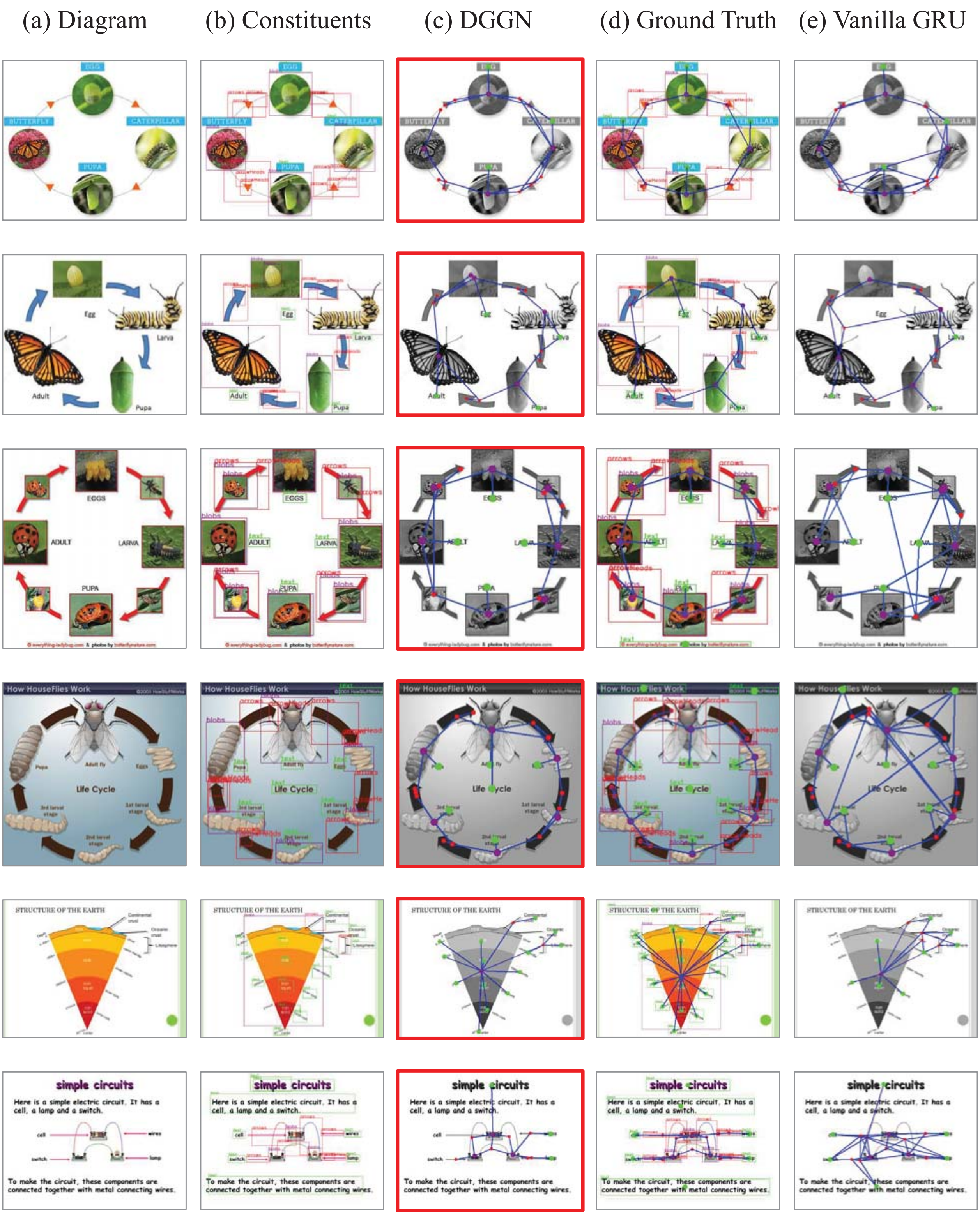

Figure 1. Additional qualitative results on diagram graph generation: (a) original diagram (b) diagram with detected constituents (c) generated graph results of DGGN (d) ground truth (e) results of baseline (vanilla GRU)

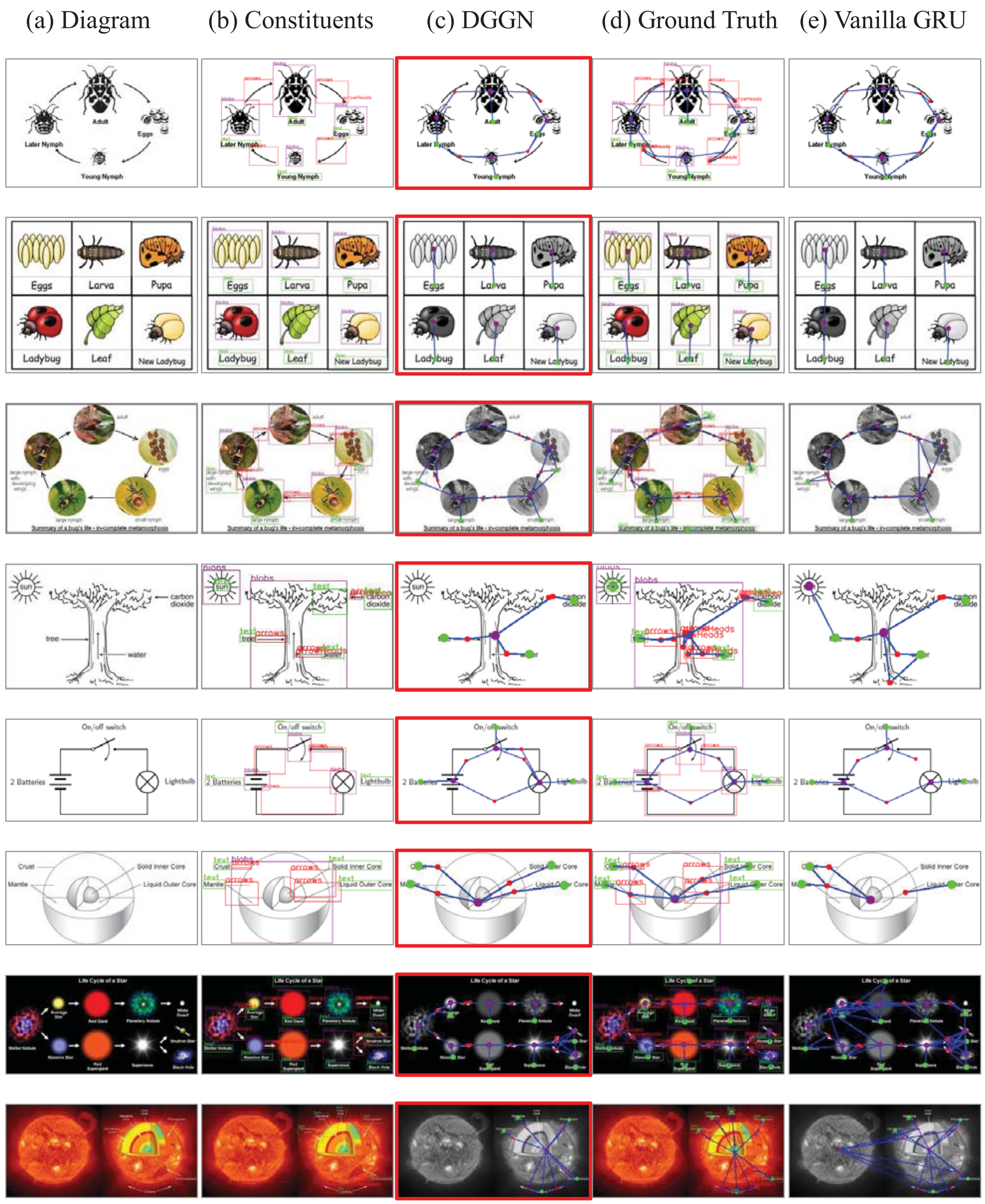

Figure 2. Additional qualitative results on diagram graph generation: (a) original diagram (b) diagram with detected constituents (c) generated graph results of DGGN (d) ground truth (e) results of baseline (vanilla GRU)

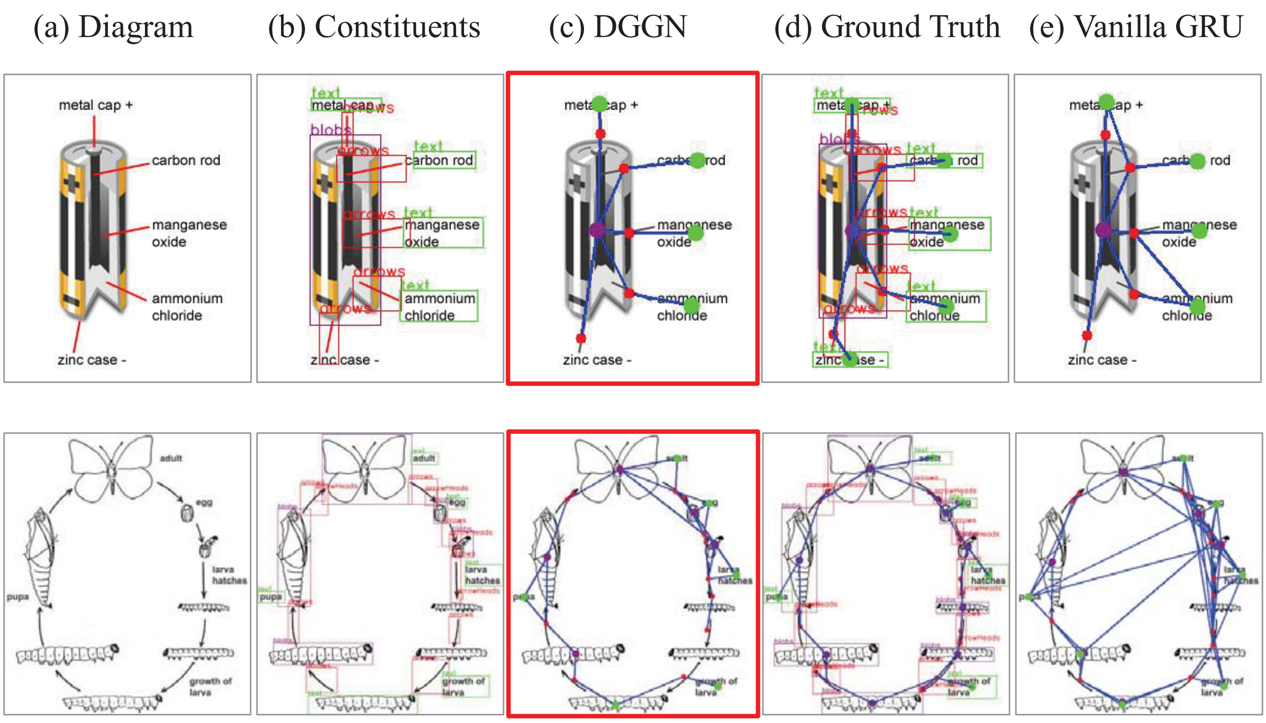

Figure 3. Additional qualitative results on diagram graph generation: (a) original diagram (b) diagram with detected constituents (c) generated graph results of DGGN (d) ground truth (e) results of baseline (vanilla GRU)

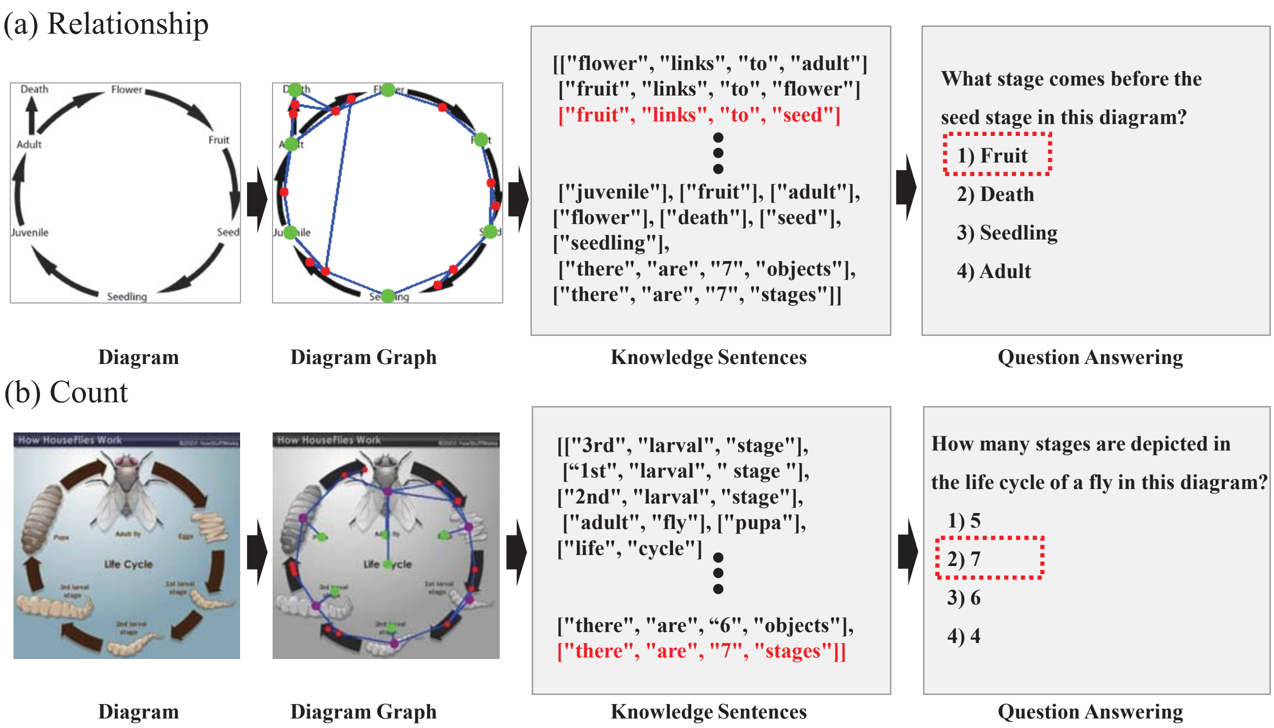

Figure 4. Additional qualitative results on question answering: (a) a question about relationship. (b) a question about the count of stages