

DEEPSOLO
ĐƯỢC HƯỚNG DẪN BỞI TỪ ĐIỂN
CHO VIỆC PHÁT HIỆN VÀ NHẬN
DẠNG VĂN BẢN TIẾNG VIỆT
TRONG CẢNH

Đoàn Nhật Sang - 21522542

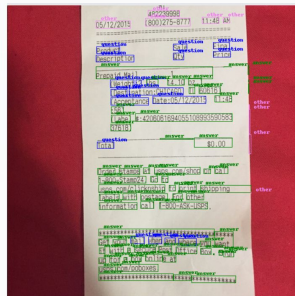
Tóm tắt

- Lớp: CS519.011
- Link Github của nhóm:
<https://github.com/dnsang1611/CS519.011>
- Link YouTube video:
<https://www.youtube.com/watch?v=MFYhRVZKsIY>



Đoàn Nhật Sang

Giới thiệu



Linear, RNN, MLP, DeConv



Generic Object Query

separately trained

(d) TTS

Linear#1, MLP



Point Query+Box Prior

(e) TESTR

Linear#2



Char Query+1D Encoding

Linear#1 & #2, MLP#1 & #2

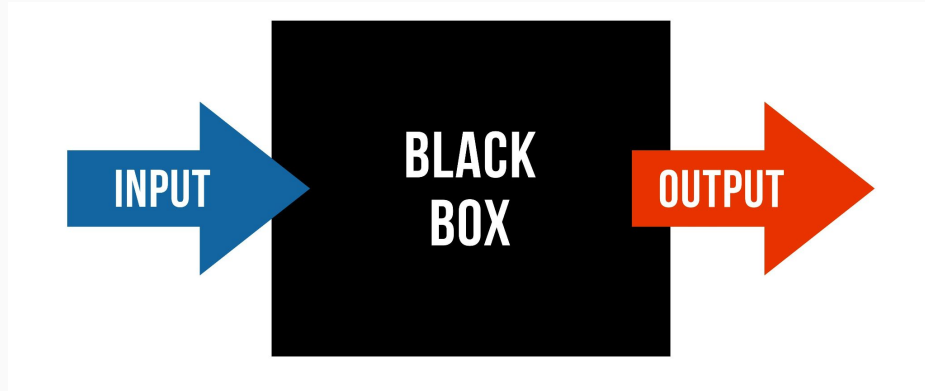


Point Query+Point Prior

(f) DeepSolo

Mục tiêu

- Xây dựng UITText
- Tìm hiểu DeepSolo
- Dictionary-guided DeepSolo



Nội dung và Phương pháp

Nội dung 1: Tìm hiểu bài toán, cách xây dựng bộ dữ liệu

- Tìm hiểu bài toán
- Precision, recall, hmean
- Xây dựng UITText:
 - 2000 ảnh - 5 người
 - CTW1500



Nội dung và Phương pháp

Nội dung 2: Tìm hiểu DeepSolo và point query

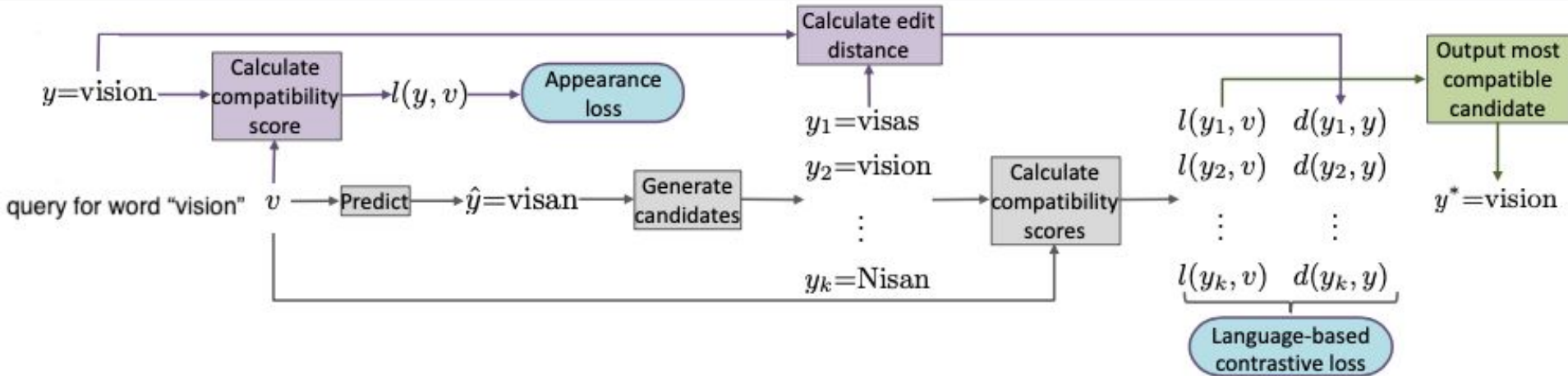
- Encoder, decoder, đầu phân loại
- Multi-scale Deformable Attention
- Factorized Self-Attention
- Positional query = $\text{MLP}(\text{PE}(\text{Coord}))$
- Tiêu chí so khớp văn bản



Nội dung và Phương pháp

Nội dung 3: Cải tiến, cài đặt, đánh giá mô hình

- Từ điển trong huấn luyện



(b) Dictionary-guided character classification head

Nội dung và Phương pháp

Nội dung 3: Cải tiến, cài đặt, đánh giá mô hình

- Từ điển trong huấn luyện
- VinText, UITText
- Nhãn đường



Kết quả dự kiến



```
31 def __init__(self, path):
32     self.file = None
33     self.fingerprints = set()
34     self.logdups = True
35     self.debug = debug
36     self.logger = logging.getLogger(__name__)
37     if path:
38         self.file = open(os.path.join(path, 'requests.json'),
39                          'w')
40         self.file.seek(0)
41         self.fingerprints.update(s.request for s in self.requests)
42
43 @classmethod
44 def from_settings(cls, settings):
45     debug = settings.getbool('SUPERLION_DEBUG')
46     return cls(job_dir(settings), debug)
47
48 def request_seen(self, request):
49     fp = self.request_fingerprint(request)
50     if fp in self.fingerprints:
51         return True
52     self.fingerprints.add(fp)
53     if self.file:
54         self.file.write(fp + os.linesep)
55
56 def request_fingerprint(self, request):
57     return request_fingerprint(request)
```

Tài liệu tham khảo

- [1]. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko: End-to-End Object Detection with Transformers. ECCV (1) 2020: 213-229
- [2]. Minghui Liao, Guan Pang, Jing Huang, Tal Hassner, Xiang Bai: Mask TextSpotter v3: Segmentation Proposal Network for Robust Scene Text Spotting. ECCV (11) 2020: 706-722
- [3]. Linjie Xing, Zhi Tian, Weilin Huang, Matthew R. Scott: Convolutional Character Networks. ICCV 2019: 9125-9135
- [4]. Yair Kittenplon, Inbal Lavi, Sharon Fogel, Yarin Bar, R. Manmatha, Pietro Perona: Towards Weakly-Supervised Text Spotting using a Multi-Task Transformer. CVPR 2022: 4594-4603
- [5]. Xiang Zhang, Yongwen Su, Subarna Tripathi, Zhuowen Tu: Text Spotting Transformers. CVPR 2022: 9509-9518
- [6]. Maoyuan Ye, Jing Zhang, Shanshan Zhao, Juhua Liu, Tongliang Liu, Bo Du, Dacheng Tao: DeepSolo: Let Transformer Decoder with Explicit Points Solo for Text Spotting. CVPR 2023: 19348-19357
- [7]. Nguyen Nguyen, Thu Nguyen, Vinh Tran, Minh-Triet Tran, Thanh Duc Ngo, Thien Huu Nguyen, Minh Hoai: Dictionary-Guided Scene Text Recognition. CVPR 2021: 7383-7392
- [8]. Yuliang Liu, Lianwen Jin, Shuaitao Zhang, Sheng Zhang: Detecting Curve Text in the Wild: New Dataset and New Solution. CoRR abs/1712.02170 (2017)