Dictionary-guided Scene Text Recognition: Supplementary material

Nguyen Nguyen[1], Thu Nguyen[1,2,4], Vinh Tran[6],
Minh-Triet Tran[3,4], Thanh Duc Ngo[2,4], Thien Huu Nguyen[1,5], Minh Hoai[1,6]

[1]VinAI Research, Hanoi, Vietnam; [2]University of Information Technology, VNU-HCM, Vietnam;
[3]University of Science, VNU-HCM, Vietnam;
[4]Vietnam National University, Ho Chi Minh City, Vietnam;
[5]University of Oregon, Eugene, OR, USA; [6]Stony Brook University, Stony Brook, NY, USA

{v.nguyennm, v.thunm15, v.thiennh4, v.hoainm}@vinai.io
thanhnd@uit.edu.vn, tmtriet@fit.hcmus.edu.vn, tquangvinh@cs.stonybrook.edu

This supplementary material contains additional qualitative results that could be included in the main paper due to page limit.

As explained in the main paper, using the naive way of using a dictionary, where the dictionary word with the smallest edit distance to the intermediate result is taken as the final output, does not always work. Fig. 1 shows two cases where this naive approach fails, while the proposed method correctly recognize the text instances in the photos.

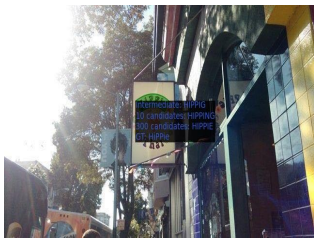
(a) Intermediate: CAFO
Closest: CFO
Proposed: CAFE
Ground-truth: CAFE

(b) Intermediate: Ylower/Tower
Closest: lower/lower
Proposed: Flower/Power
Ground-truth: Flower/Power

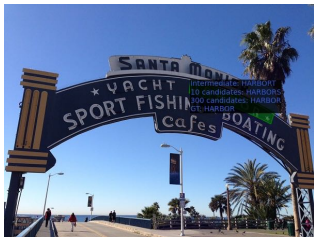
Figure 1: Several cases where our proposed approach of using the dictionary leads to correct recognition performance, while the naive approach of casting the intermediate result to a dictionary word fails.

As reported in the main paper, the performance the proposed method ABCNet+D increases as the number of candidates considered during inference increases. This can be attributed to its compatibility scoring model to calculate the compatibility score between the visual feature and the candidates. Fig. 2 shows several cases where having the candidate set of size 10 is not enough, but the text instances are correctly recognized when we increase the size of the candidate sets to 300.

(a) Intermediate: PETTT
10 candidates: PETTY
300 candidates: PETIT
Ground-truth: PETIT



(b) Intermediate: HIPPlG
10 candidates: HIPPING
300 candidates: HIPPIE
Ground-truth: HIPPIE



(c) Intermediate: HARBORT
10 candidates: HARBORS
300 candidates: HARBOR
Ground-truth: HARBOR

(d) Intermediate: TOWAV
10 candidates: TWAT
300 candidates: COWAY
Ground-truth: COWAY

Figure 2: The accuracy of the proposed method correlates with the size of the candidate set considered during inference. As shown in the main paper, using the candidate set size 10 improves the recognition performance significantly. However, sometimes, 10 might not be enough; the model can correctly recognize the text instances with 300 candidates but not 10.

The remaining set of figures display representative detection and recognition results of ABCNet+D on several randomly chosen images from the datasets considered in this paper.


GEMBIRA
LOKA


MELBOURNE
AUSTRALIA
TAJ MAHAL
INDIA


HAIGHT
ASHBURY
VINTAGE
Vintage


CANNERY
ROW
DIRECTORY

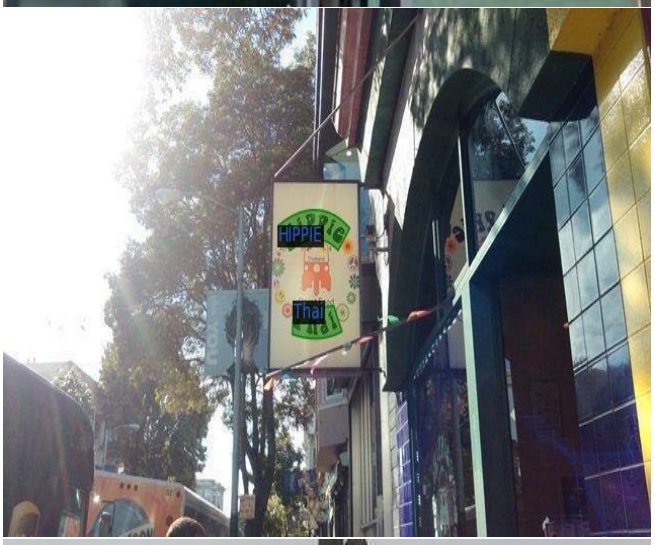

BOUDIN
1849
SINCE
SOURDOUGH
BAKERY
CAFE


LOUIE
LINGUINI'S
SEAFOOD
SHACK


TC
TOUS


THROWDOWN
THURSDAYS
DINN
AUG
4TH
10P
NOW ON
THROWDOWN


NATURE VALLEY
Crunchy Granola Bars
Oats & Honey


HIPPIE
Thai


Kellogg's
KRISPIES
ICE
TREATS
THE
Original
TREATS
SQUARES
MARSHMALLOW
ERY


WHARF
FISHERMANS
SAN
FRANCIS


SAN FRANCISCO
SF
49ERS
49ERS


Square
Chirardelli


LOUIE THE PENGUIN'S
PASTA • SEAFOOD • PIZZA
UPSTAIRS
CANNERY


RIDGWELL WINDOWS
CONSERVATORIES
DOORS WINDOWS CONSERVATORIES
UPVC FASCIA PORCHES SOFFITS & GUTTERS
www.ridgwellwindows.co.uk


MONTEREY
FIRE


MORNING
CALL
COFFEE
STAND


PLANET
HOLLYWOOD


Madly
British


ALL INDIA PERMIT
TOURIST VEHICLE


TEA HOUSE
WINE TASTING
GASTRONOMY


CHIPOTLE
MEXICAN GRILL


chucuk-chucuk
restaurant


QUESNEL
GOLD
PAN
City


SANTA
MONICA
MACHT
HARBOR
FISHING
SPORT
BOATING
Cafes


ERECTED


Turkish
Delight
Welcom


BAZAAR
GRAND
Club

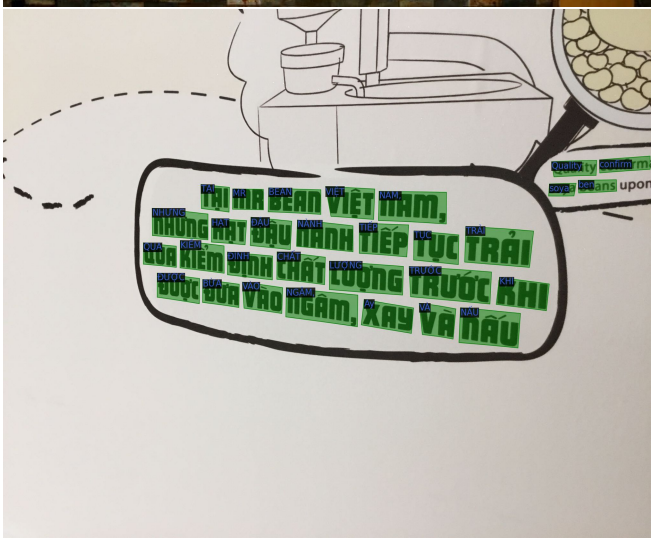

SANRIO
HELLO KITTY TOWN


XEM PHIM
XUYÊN
ĐÊM
CGV
85,000
NƯỚC
BẮP
CÀN MẠNH


TAN
ĐÀ
SPA
RESORT


THÀNH PHỐ HỒ CHÍ MINH
QUYẾT TÂM ĐƯA NGHỊ QUYẾT
ĐẠI HỘI XI CỦA ĐẢNG ĐI VÀO CUỘC SỐNG


Lipton
TRÀ XANH
Thanh Sạch cơ thể
ngay hôm nay


THỊ HIẾU BÌNH VIỆT NAM,
NHỮNG HẠT ĐẬU MẠNH TIẾP TỤC TRÀI
CỦA KIÊN ĐỊNH CHẤT LƯỢNG TRƯỚC KHI
ĐƯỢC ĐƯA VÀO NGÂM, XAY VÀ NẤU


MOTO
KIỂU
PAINT
VĨNH
NGUYÊN 2
BẠN SƠN DẦN SƠN NƯỚC SƠN THỢ MỘC ĐIỆN
ỐNG NƯỚC BỊ ĐÁ MÀI ĐÁ CẮT ĐUA HẠN


Càphê
NHA
SAN


NHA KHOA 108
PHÒNG KHÁM BỆNH TƯ NHÂN CHUYÊN KHOA
RĂNG - HÀM - MẶT
KHÁM,TƯ VẤN MIỄN


XIN ĐỪNG GÁC CHÂN
Please Do Not Keep your Legs
mobifone
WIFI
3G
Free


POZAA TEA
TRÀ SỮA ĐÀI LOAN


Coi Quán
BA THẰNG HÙNG
GIỮ XE
MIỄN PHÍ
TRẢ TIỀN
TRÁCH NHIỆM


BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NGOẠI THƯƠNG
CHÀO MỪNG TÂN SINH VIÊN KHÓA 57
Hà Nội, tháng 08 năm 2018

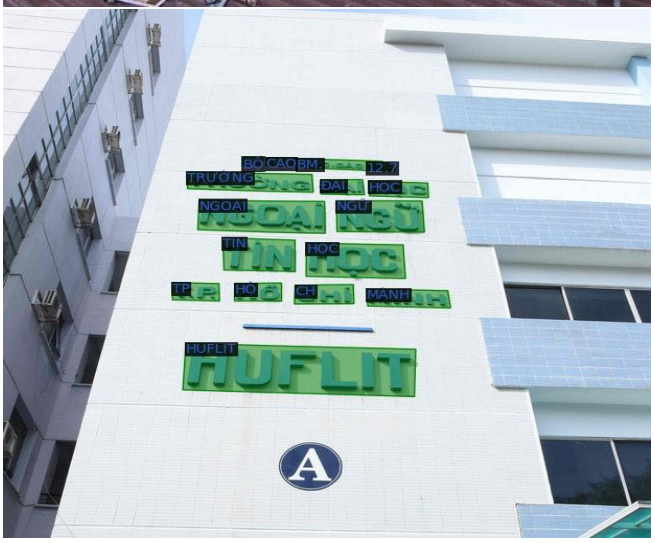

KHUYẾN MẠI
BÁNH COSY QUẾ CÁC VỊ
132G
11.700
10.500
3%
25/01/2020-29/02/2020
COSY


KHU PHỐ VĂN HÓA
LÊ HỒNG PHONG


TỔNG KẾT CUỐI NĂM - THANH LÝ TOÀN BỘ
XẢ HẾT - LỖ CUNG BẢN
GIÁ


thế giới điện thoại di động
TIẾN MOBILE
ĐIỆN THOẠI - PHỤ KIỆN - SỬA CHỮA


tiệm trà chanh
1975


TRƯỜNG ĐẠI HỌC
NGOẠI NGỮ
TIN HỌC
HUFLIT


NHÀ
THUỐC


BIA HÀ NỘI
LÀNG SM


TRUNG TÂM
NGOẠI NGỮ


THEP
Pomina


SAIGON
HỒNG HẠNH


Bánh MÌ & Chè CALI
BAKERY


BẢO TRÌ
THANG MÁY

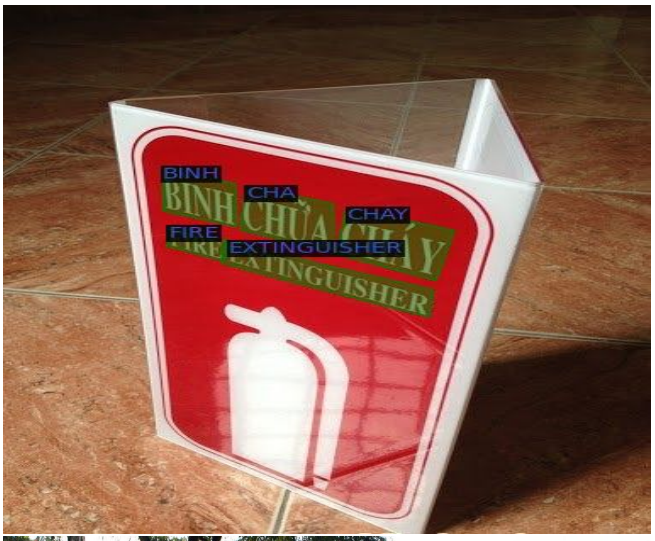

TIẾP NHẬN HỒ SƠ
ĐĂNG KÍ TỜ KHAI


CÔNG VIÊN
THÀNH PHỐ VĨNH LONG


TRUNG TÂM THỂ THAO QUẬN 1
THỂ DỤC


TrangAn
Belgi


CHÈ - TRÀ SỮA ĐÀ LẠT
"Độc quyền"
(0988 291 391)

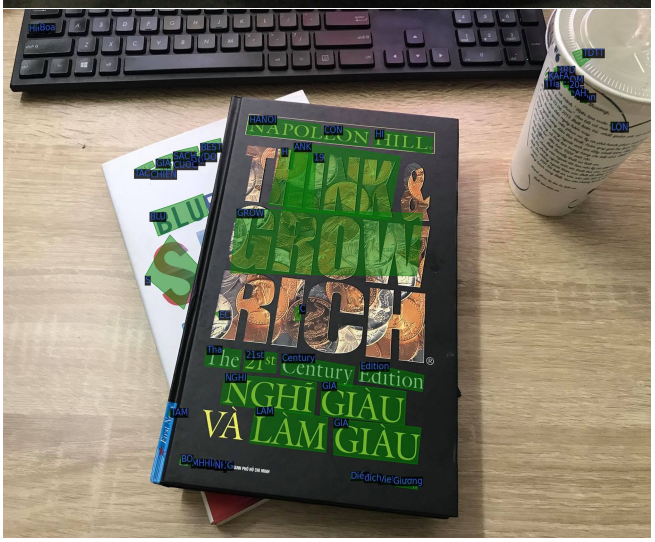

BINH
CHA
CHAY
FIRE
EXTINGUISHER


ĐỒNG VĂN
KINH CHÀO QUÝ KHÁCH
TẠM BIỆT QUÝ KHÁCH
HẸN GẶP LẠI
Good luck! See you again


60 NĂM
10.10.1954


HỌC VIỆN CHÍNH TRỊ CÔNG AN NHÂN DÂN
LỚP CAO CẤP CHÍNH TRỊ K5
THAM QUAN THỰC TẾ GIA LAI - BMT - ĐẮK LẮK 28-2/4/2017
29C-011.38
01138


00:42:28
Kính chào quý khách


NAPOLEON HILL
THINK & GROW RICH
The 21st Century Edition
NGHĨ GIÀU VÀ LÀM GIÀU