

트랜슬레이션 임베딩 기반 관계 학습을 이용한 GUI 위젯 인식

Recognition of GUI Widgets Utilizing Translational Embeddings based on Relational Learning

박 민 수*, 석 호 식*

Min-Su Park*, Ho-Sik Seok*

Abstract

CNN based object recognitions have reported splendid results. However, the recognition of mobile apps raises an interesting challenge that recognition performance of similar widgets is not consistent. In order to improve the performance, we propose a noble method utilizing relations between input widgets. The recognition process flows from the Faster R-CNN based recognition to enhancement using a relation recognizer. The relations are represented as vector translation between objects in a relation space. Experiments on 323 apps show that our method significantly enhances the Faster R-CNN only approach.

요 약

CNN 기반의 객체 인식 성능은 매우 우수한 것으로 보고되고 있지만 모바일 기기의 앱 GUI와 같이 일반적으로 생각하기에 잡음이 적고 분명하게 인식될 수 있을 것으로 기대되는 환경에 적용해보면 인간의 관점에서 매우 유사한 GUI 입력 위젯들이 의외로 잘 인식되지 않는다는 문제가 발생한다. 본 논문에서는 CNN의 입력 위젯 인식 성능을 향상시키기 위하여 모바일 앱의 GUI를 구성하는 객체들의 관계를 활용하는 방법을 제안한다. 제안 방법에서는 (1) CNN 기반의 객체 인식 도구인 Faster R-CNN을 이용하여 모바일 앱을 구성하는 입력 위젯을 1차 인식한 후 (2) 위젯 인식을 향상을 위하여 객체 간의 관계를 활용하는 방법을 결합하였다. 객체 간의 관계는 표현 공간상에서의 벡터의 평행 이동을 활용하여 표현되었으며, 총 323개의 앱에서 생성한 데이터에 적용한 결과 Faster R-CNN만을 사용한 경우와 비교하여 위젯 인식률을 상당히 개선할 수 있음을 확인하였다.

Key words : Deep neural nets, Object recognition, Relation learning, Widget recognition, App test

* Dept. of Computer and Communication Engineering,
Kangwon National University

★ Corresponding author

E-mail : hsseok@kangwon.ac.kr, Tel : +82-33-250-6381

※ Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2016R1D1A1B03931615) and 2017 Research Grant from Kangwon National University.

Manuscript received Sep. 5, 2018; revised Sep. 11, 2018; accepted Sep. 13, 2018

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

CNN(Convolutional neural network)에 기반한 객체 인식은 기존의 객체 인식 방법에 비교하여 놀라운 성능 향상을 보이고 있으며[1]-[3], 그 결과 다양한 애플리케이션에 적용되어 실생활에서 사용되고 있다. 하지만 모바일 앱의 GUI(Graphical user interface)를 구성하는 입력 위젯(Input widget, 이하 “위젯”)을 인식하기 위하여 CNN에 기반한 객체 인식 방법을 적용하면 인식 결과가 일관적이지 못하다는 다소 놀라운 결과를 관찰하게 된다. 예를 들어 아이디나 암호를 입력하는 텍스트 입력창인 에디트텍스트(EditText)의 인식을 시도할 경우 전

혀 잡음이 존재하지 않고 색상이 뚜렷하게 구분되는 GUI 환경에서, 일부 에디트텍스트 위젯이 인식되지 않는 현상이 발생한다. 본 논문에서는 일관적이지 않은 인식이라는 문제를 해결하기 위하여 “관계”라는 정보를 부가하여 객체 인식 성능을 향상시키는 방법을 제안한다. 제안 방법은 인식하려는 객체 간의 관계를 표현 공간상에서의 평행 이동(Translation)으로 표현하는 접근법을 응용한 것으로, 주어진 문제를 새로운 표현 공간에 매핑한 후 정보를 추가하여 인식 성능을 향상시키려는 시도이다.

제안 방법은 특히 모바일 앱의 자동 테스트를 위한 GUI 화면 인식에 적용되었는데, 모바일 앱의 경우 소스 코드를 분석하면 화면 구성을 용이하게 파악할 수 있으리라 짐작하기 쉽지만, (1) UI (User interface) 구성 방식의 비표준화, (2) HTML, CSS (Cascading style sheets) 및 각종 스크립트 언어의 혼용, (3) 웹 언어의 놀라운 표현성이라는 요소가 결합하면서 XML 코드의 파싱만으로는 화면 구성을 파악할 수 없다는 문제가 발생한다. 이와 같은 문제를 해결하기 위하여 본 논문에서는 딥뉴럴넷(Deep neural nets)을 이용하여 UI 이미지로부터 위젯을 인식하는 방법을 제안하였으며, 323개의 모바일 앱 GUI에서 획득한 412장의 위젯 이미지 인식을 통하여 제안 방법이 입력 위젯 인식률을 상당히 개선할 수 있음을 확인하였다.

II. 관련연구

1. 관계 및 객체 인식 접근방법

이미지를 구성하는 객체 간의 관계가 추가되면 이미지를 보다 다양한 관점에서 이해할 수 있다. 왜냐하면 이미지를 구성하는 객체들을 인식한 후 객체 간의 관계를 추정하면 이미지의 총체적인 의미를 파악할 가능성이 높아지기 때문이다. 그래피컬 모델(Graphical model)은 객체의 관계를 표현하기에 매우 자연스러운 접근법 중 하나이다. 한 예로 “Scene graph” 모델에서는 객체, 객체의 속성, 객체 간의 관계를 CRF(Conditional random field) 모델로 표현하였다[4]. CRF를 사용하게 되면 객체의 속성과 관계가 모두 반영되기 때문에 이미지의 내용에 기반한 정보 추출이 가능하지만, 수작업으로 추출된 관계와 속성에 의존하므로 확장성에 문제가 존재한다. CoLA 또한 CRF 모델을 이용하여 의

미 컨텍스트(Semantic context)를 표현하려고 하였으나 객체 간의 관계보다는 객체 레이블링에 치중하였으며 학습된 관계의 수가 위치 관계 네 가지로 국한되었다 [5]. [6]에서는 계층 모델을 이용하여 객체 간의 의미적 관계와 기하학적 관계를 표현하였다. 객체 간의 관계를 묘사하는 언어 문장이 함께 제공될 경우 관계 학습에 큰 도움이 될 수 있다. “Visual relationship prediction model”에서는 $\langle \text{obj}_1\text{-predicate-obj}_2 \rangle$ 의 형태로 관계를 추정하기 위하여 유사한 관계는 서로 근접할 수 있도록 임베딩 공간(Embedding space)으로 프로젝션하는 방법을 사용하였다[7]. [8]에서는 동작 인식에 집중하였으며 가능한 동작의 스코어를 계산하여 동작을 예측하는 일종의 동작 분류 접근법을 취하였다. [9]에서도 동작 분류 과정에서 자연어 문장에 포함된 정보를 활용하는 접근법을 동시에 활용하는 접근법을 소개하였다. [10]에서는 DR-Net(Deep Relational Network)을 이용하여 이미지에 존재하는 객체 간의 관계를 추출하는데, CRF를 확장하여 관계 추정에 필요한 항목들의 확률값을 계산하는 딥네트워크 모델을 소개하였다.

딥뉴럴넷을 이용하여 데이터에 내재된 관계를 학습하려는 연구도 많이 시도되었다. [11]에서는 MRF(Markov random field)를 딥러닝과 결합하여 출력 변수 간의 의존도를 표현하는 방법을 제안하였다. [12]은 링크 구조나 노드의 특성을 이용하여 링크를 예측해왔던 방식과 다르게, PoG (Product of Gaussians) 구조를 사용하여 다양한 변수에 대한 추론을 연계하였으며 그 결과 링크 인디케이터와 노드 속성을 모두 고려한 예측이 가능한 모델을 소개하였다. 전이 학습 혹은 다중 임무(Multiple task) 학습에서도 관계의 학습이 중요하다. [13]에서는 태스크 간의 관계를 명확하게 모델링하지 않는다는 기존 딥러닝 계열 접근법의 한계를 극복하기 위하여 파라미터 텐서에 존재하는 임무 간 관계를 모델링할 수 있는 사전 확률 분포를 활용하였다.

주어진 이미지 내에 존재하는 객체 인식 과정에서 딥뉴럴넷은 놀라운 성능을 자랑하고 있다. 이미지 세그멘테이션 방법[14]을 적용하여 분할된 이미지에 CNN에 기반한 객체 인식 방법을 적용하면, CNN을 사용하지 않은 다른 방법보다 월등히 뛰어난 성능으로 객체를 인식할 수 있음이 보고되었다[1]-[3].

2. 모바일 앱 자동검증

최대한 빠른 시간 내에 앱을 출시하는 것이 중요해지면서 인간 테스터가 앱의 동작을 검증하지 않고 알고리즘으로 동작을 검증하는 자동 테스터가 주목받고 있다(자동 테스트 구현에 대한 보다 자세한 소개는 [15] 참조). 모바일 앱 더 나아가 프로그램을 자동 테스트하려면 앱의 동작 모델을 확보하는 것이 가장 확실하지만, 앱의 동작 기술 모델을 작성하는 것 자체가 매우 방대한 작업이어서 현실적으로 용이하지 않으므로 무작위로 동작을 생성한 후 후속 동작을 관찰하는 방법이 의외로 효율적인 방법으로 활용되고 있다. 그러나 앱의 기능을 활용하기 위해서는 주어진 GUI 입력에서 활용 가능한 위젯에 해당하는 부분을 인식해야 하는데 (1) UI 구성 방식의 비표준화, (2) HTML, CSS 및 각종 스크립트 언어의 혼용, (3) 웹 언어의 놀라운 표현성이라는 요소가 결합하면서 앱을 구성하는 위젯을 XML 코드로부터 추출하기가 매우 어렵다. 본 논문에서는 XML 코드를 활용하지 않고, 인간이 인식하는 방법과 같이 주어진 UI 화면으로부터 사용 가능한 위젯을 인식하는 방법을 소개하고자 한다.

3. 임베딩기반 관계인식

객체 간의 관계를 학습하기 위하여 제안된 방법 중 평행 이동을 활용한 방법이 최근 소개되고 있다. [16]에서는 여러 개의 엔터티와 관계를 저차원의 벡터 스페이스에 임베딩(Embedding)하는 방법인 TransE를 소개하였는데, TransE는 엔터티의 저차원 임베딩에서의 평행 이동으로 관계를 해석하는 흥미로운 접근법이었다. VTransE[17]는 TransE를 이미지 데이터에 존재하는 관계 인식에 활용할 수 있도록 확장한 방법으로, 객체 인식과 관계 분류를 위한 기능을 구현하여 이미지 데이터에의 적용 가능성을 입증하였다. 보다 구체적으로 개체 간의 subject-predicate-object 관계를 TransE에서는 저차원 벡터 연산인 $\vec{s} + \vec{p} \approx \vec{o}$ 로 표현하는데, VTransE에서는 관계 평행 이동 벡터 외에 프로젝션 행렬을 추가 학습한 후 $\vec{W}_s \vec{x}_s + \vec{t} \approx \vec{W}_o \vec{x}_o$ 의 형태로 관계를 표현한다.

본 논문에서는 기본적으로 Faster R-CNN으로 대표되는 CNN 기술과 VTransE의 관계 인식 학습을 결합하여 위젯 인식 성능을 높이고자 하였다. 또한 모바일 앱의 GUI 위젯 인식 성능을 향상시킬

수 있도록 객체 인식과 관계 인식의 결합과정에서 탐색 공간의 수를 줄이고 탐색 성능을 높이기 위하여 모바일 앱의 GUI에 고유한 제약 조건을 부가하여 위젯 인식률을 향상시키는 방법을 제안한다.

III. 이미지 인식과 관계 학습

1. 객체 및 관계 인식에 기반한 위젯 인식

위젯 인식을 위하여 본 논문에서 제안된 알고리즘은 기본적으로 Faster R-CNN[3]을 이용한 1차 위젯 인식, 1차 인식된 결과의 가공, 그리고 VTransE를 활용한 위젯의 종류 확정의 3단계로 구성된다(알고리즘 1). 입력으로는 별도의 가공을 거치지 않은 모바일 앱의 UI 이미지가 주어지며, 출력으로는 주어진 입력 이미지에 존재하는 위젯에 대하여 해당 위젯의 인식 결과(위젯의 카테고리 및 좌표 정보)가 생성된다.

Algorithm 1. Widget recognition.

Input UI images of a mobile app

- 1: 1st widget detection step through Faster R-CNN.
- 2: Generation of pairs of detected widgets.
- 3: 2nd widget detection step considering relations.

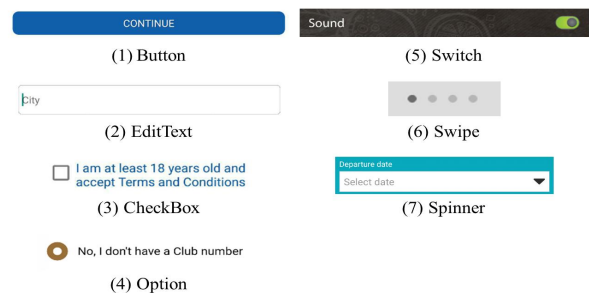


Fig. 1. Widget categories.

그림 1. 위젯 카테고리

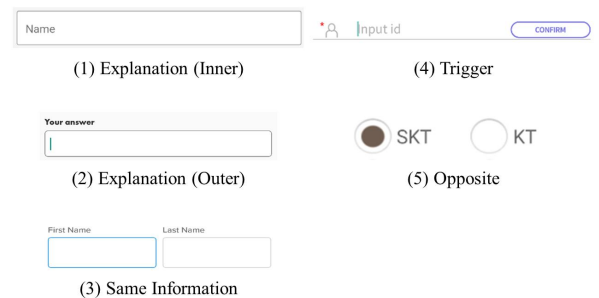


Fig. 2. Relationship categories.

그림 2. 관계 카테고리

가. Faster R-CNN을 활용한 1차 위젯 인식

1단계에서는 Faster R-CNN으로 입력된 이미지에 존재하는 위젯들의 1차 인식을 시도한다. 인식 작업을 위하여 안드로이드 앱스토어에서 총 205개의 모바일 앱을 다운받아 버튼(Button), 에디트텍스트, 체크박스(Checkbox), 옵션(Option), 스와이프(Swipe), 스위치(Switch), 스피너(Spinner), 가이드 텍스트(Guiding text)의 총 8가지 카테고리 레이블을 부착한 훈련 이미지 2631개를 생성한 후 Faster R-CNN 훈련에 활용하였다(위젯의 모습은 그림 1과 그림 2 참조). 다양한 위젯 이미지를 인식하도록 훈련된 Faster R-CNN은 주어진 모바일 앱 이미지에 대하여, 각 이미지 영역이 어떤 위젯 카테고리에 속할 것인지 여부를 추정하여 각 영역에 대한 (1) 좌표 정보 (2) 해당 영역의 위젯 카테고리 정보 (3) 특정 카테고리에 속할 스코어 정보를 반환한다.

나. 위젯 쌍 생성

2단계에서는 1단계에서 생성된 부가 정보를 활용하여 위젯의 쌍을 생성한다. 1단계에서 화면 분할이 이루어진 상태이므로 분할된 특정 영역을 선택한 후, 해당 영역의 카테고리 정보와 다른 영역의 카테고리 정보를 결합하여 2개 위젯 영역의 쌍을 생성한다. 구체적으로 생성되는 정보는($obj_{coord}^1, obj_{category}^1, obj_{coord}^2, obj_{category}^2$)와 같이 영역 1의 좌표(obj_{coord}^1)와 영역 1의 카테고리($obj_{category}^1$), 영역 2의 좌표(obj_{coord}^2)와 영역 2의 카테고리($obj_{category}^2$)가 결합된 벡터의 형태를 취하게 된다.

다. 관계를 고려한 위젯 식별

2단계에서 생성된 벡터를 VTransE에 입력하게 되면 해당 벡터가 갖는 관계를 출력하며, 그 관계를 반영하여 각 영역에 해당하는 위젯들의 최종 카테고리를 결정한다. 이를 위하여 먼저 VTransE를 학습시킬 필요가 있다. 우리는 총 245개의 앱*에서 획득한 677개의 이미지에서 (1) 가이드텍스트-위젯(외부 설명 관계), (2) 가이드텍스트-위젯(내부 설명 관계), (3) 에디트텍스트-스피너 및 에디트텍스트(동일정보 관계), (4) 버튼-에디트텍스트(트리거 관계), (5) 옵션-옵션(반대 관계)의 관계를 추출하

* Fast R-CNN과 달리 관계 데이터는 다양한 형태의 위젯 쌍이 필요하므로 Faster R-CNN 훈련에 활용한 205개의 애플리케이션에 40개의 애플리케이션을 추가하여 사용함

여 VTransE를 학습시켰다. 2단계에서 생성된 벡터는 모든 위젯에 대해 쌍을 만드므로 입력된 화면 이미지에 존재하는 모든 후보 위젯 영역 N개에 대하여 $N \times C_2$ 개의 쌍을 생성한 후 VTransE를 수행하여야 한다. 이 경우 전혀 관계가 없는 쌍에 대해서도 VTransE를 수행하기 때문에 시간 및 결과 관점에서 보았을 때 비효율적이다. 따라서 생성된 모든 후보 쌍에 대하여 VTransE를 수행하지 않고 다음에 설명하는 제약조건을 활용하여 VTransE의 연산 대상을 축소하였다: (1) 분류한 5개의 관계는 가이드 텍스트와 버튼, 에디트텍스트와 스피너 같은 특정한 위젯 쌍에만 관계를 맺는다. 따라서 단계 2에서 생성된 쌍을 보았을 때 각 위젯의 카테고리에 대한 적합한 관계가 존재하지 않으면 VTransE를 수행하지 않도록 하였다. (2) 일반적으로 모든 위젯은 높이보다 너비가 크다는 점에 착안하여 높이가 너비보다 큰 후보 영역에 높이가 너비보다 1.5배 이상 클 경우 VTransE를 수행하지 않도록 하였다. (3) 2단계에서 스코어가 높은 위젯들의 평균 높이를 구한 후 입력으로 들어온 위젯이 평균 높이보다 1.5배 크거나 0.5배보다 작을 경우 VTransE를 수행하지 않도록 하였다. 위와 같은 제약조건을 통해 관계 쌍의 수를 줄인 벡터 집합에 훈련된 VTransE 모델을 적용하면, 해당 쌍에 대해 5개의 관계 카테고리 중 하나를 찾고 해당 관계에 대한 스코어를 함께 반환한다. 그러나 VTransE를 거쳐 반환된 관계 카테고리가 에디트텍스트와 스피너를 트리거 관계로 판단하는 등의 부정확한 결과가 관찰되었다. 이런 문제를 해결하기 위하여 반환된 카테고리 와 스코어를 기반으로 (1) 내부 설명 관계라고 판단하였으나 가이드 텍스트가 위젯 외부에 존재하는 경우, (2) 내부 설명 관계가 아닌 다른 관계를 갖는다고 판단한 위젯들이 각 위젯의 내부에 위치하는 경우, (3) 관계를 갖는 위젯 중 하나를 선택하여 해당 위젯의 위아래로 해당 위젯의 높이만큼의 간격을 추가한 후, 쌍을 이루는 다른 위젯이 해당 간격 안에 포함되지 않는 경우를 결과에서 제외하는 등의 제약조건을 추가로 적용하였다. 마지막으로 제약조건을 통과하여 인식된 관계 카테고리 와 관계 스코어 정보를 기존 영역 정보와 합해 반환하면 해당 위젯 쌍이 갖는 관계 카테고리를 판단할 수 있으며, 이를 기반으로 Faster R-CNN에서 인식하지 못했던 위젯을 인식할 수 있게 된다.

IV. 실험결과

1. 테스트 데이터 생성

제안 방법의 성능을 확인하기 위하여 안드로이드 앱스토어에서 다운 받은 아마존, 트위터, 구글과 같은 323개의 상용 애플리케이션에서 총 412개의 이미지를 추출하여 성능 평가를 수행하였다. Faster R-CNN과 VTransE의 훈련 과정에서 사용한 애플리케이션에 83개의 애플리케이션을 추가하였으며, 훈련 과정에서 사용하였던 화면의 이미지들이 아닌 화면의 이미지로부터 테스트 데이터를 생성하였다. 각각의 이미지들은 평균적으로 7개의 위젯을 포함하고 있으며, 11개의 관계 쌍을 가지고 있다.

2. 인식 결과 및 분석

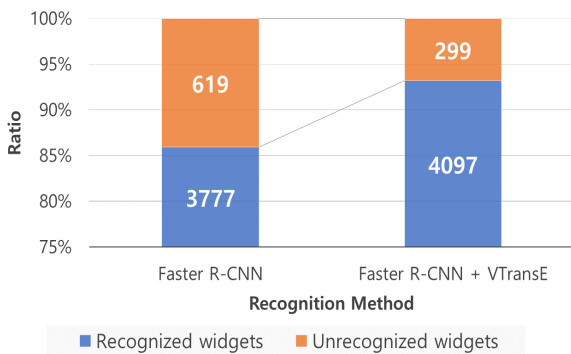


Fig. 3. Recognition performance: without VTransE vs With VTransE.

그림 3. VTransE를 추가하지 않았을 때와 추가하였을 때 인식 성능 비교

자체 생성한 테스트 데이터에 대한 위젯 인식 결과를 그림 3에 표시하였다. 성능 평가를 위하여 Faster R-CNN을 이용한 위젯 인식 성능과 Faster R-CNN+VTransE를 사용하였을 때 위젯 인식 결과를 비교하였다. VTransE를 적용하지 않은 경우, 즉 Faster R-CNN만을 사용했을 경우 위젯의 인식률은 85.9%이지만 VTransE과 함께 사용했을 경우 인식률이 93.2%로 향상되었다. 그림 4~6에서는 위젯 카테고리별로 제안 방법의 인식률 향상을 비교하였다. 버튼의 경우 인식률이 87.6%에서 97.3%로 향상되었고, 에디트텍스트는 내부 설명 관계의 경우 80.1%의 인식률을 보였고, 외부 설명 관계의 경우 89.4%의 인식률을 보였으나 관계 정보를 추가하면 각각 90.5%와 92.6%로 향상되었다. 체크박스는 인식률이 83.8%에서 93.7%로 향상되었으며, 옵션

은 73.8%에서 82.3%로 향상되었다. 마지막으로 스피너는 내부 설명 관계일 경우 92.9%에서 94.1%로 인식률이 다소 상승했으며 외부 설명 관계인 스피너와 스위치는 인식률 변화에 큰 차이가 없었다. 설명 관계 이외의 동일정보 관계, 트리거 관계, 반대 관계의 경우 위젯 간의 관계를 통하여 주변에 있는 위젯을 찾을 것으로 예상하였으나, 동일정보 관계 및 트리거 관계의 경우 인식 성능 향상이 관찰되지 못하였고, 반대 관계는 기존 90.7%에서 97.3%로 향상되었다.

그림 4~6의 카테고리별 인식 성능을 보면 버튼, 에디트텍스트, 체크 박스, 옵션의 경우 “관계” 정보를 추가했을 때 인식률이 향상되며, 그 외 카테고리의 경우 “관계” 정보의 부가가 인식률 향상에 거의 도움이 되지 않음을 알 수 있다.

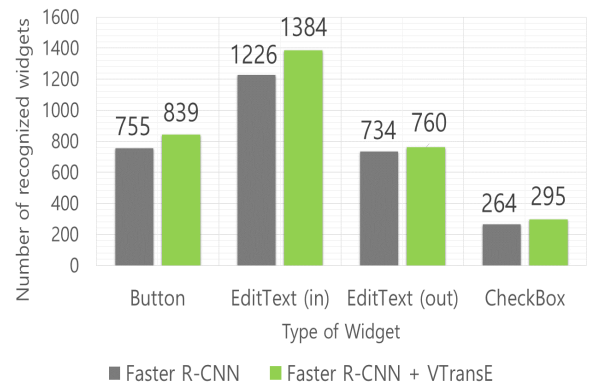


Fig. 4. Recognition performance in detail (Button, EditText, CheckBox).

그림 4. 위젯 카테고리별 인식 성능 비교(버튼, 에디트텍스트, 체크박스)

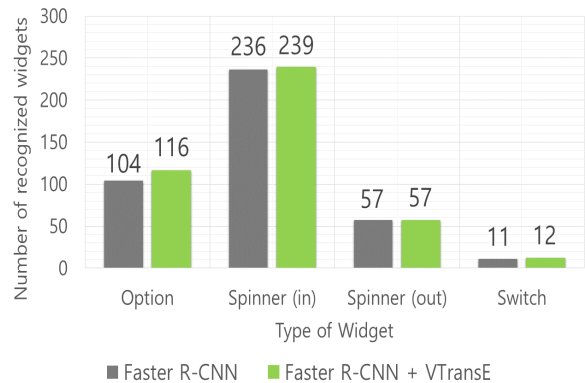


Fig. 5. Recognition performance in detail (Option, Spinner, Switch).

그림 5. 위젯 카테고리별 인식 성능 비교(옵션, 스피너, 스위치)

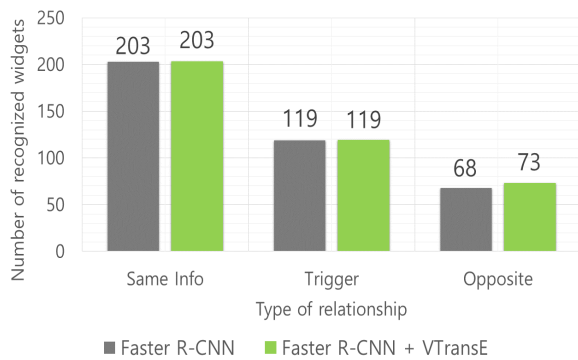


Fig. 6. Recognition performance in detail (Same info, Trigger, Opposite).

그림 6. 위젯 카테고리별 인식 성능 비교(동일정보, 트리거, 반대 관계)

이것은 학습에 사용한 5가지 관계((1) 가이드텍스트-위젯(외부 설명 관계), (2) 가이드텍스트-위젯(내부 설명 관계), (3) 에디트텍스트-스피너 및 에디트텍스트(동일정보 관계), (4) 버튼-에디트텍스트(트리거 관계), (5) 옵션-옵션(반대 관계))과 밀접한 관계가 있다. 일반적인 모바일 UI에서 위젯의 활용 모습을 보면 에디트 텍스트와 버튼, 체크 박스는 해당 위젯에 입력해야 하는 정보를 설명하는 간략한 정보와 함께 등장하는 경우가 많다. 하지만 스피너나 옵션의 경우 별도의 가이드텍스트가 없어도 해당 위젯을 클릭했을 때 등장하는 선택 메뉴를 통해 해당 위젯의 기능을 파악할 수 있도록 설계되어 있다. 따라서 버튼, 에디트텍스트, 체크 박스는 다른 위젯과의 관계를 활용하여 인식률을 높이는 것이 가능하지만 스피너와 옵션에서는 관계 활용이 인식 성능 향상에 도움이 되지 않는다. 그림 6에 표시된 경우도 마찬가지로 동일 및 반대 관계는 UI의 구성에 따라 성립하지 않을 수도 있는 관계이기 때문에 인식률 향상에 도움이 되지 못하며 트리거 관계의 경우 트리거에 사용되는 버튼의 위치와 개수가 근접하지 않은 경우가 많고 시각 정보만으로 버튼을 정확하게 식별하는 것이 어려운 경우가 많기 때문에 현재 고려한 제약 조건만으로는 인식률을 향상시키는데 어려움이 있었다. 또한 이는 Faster R-CNN에서 결과로 도출된 스코어와도 밀접하게 관련되어 있다. Faster R-CNN에서 생성된 결과는 크게 스코어가 큰 위젯과 스코어가 작은 위젯으로 나눌 수 있다. VTransE를 수행하기 위해서 각각의 위젯들에 대한 쌍을 생성할 때 높은 스코어-높은 스코어, 높은 스코어-낮은 스코어, 낮

은 스코어-낮은 스코어의 쌍이 생성된다. 높은 스코어-높은 스코어 쌍의 경우 Faster R-CNN 결과에서 찾을 수 있어 성능 향상에는 도움이 되지 않았으며, 낮은 스코어-낮은 스코어 쌍의 경우 불필요한 영역을 가리키기 때문에 관계 스코어가 낮아 최종 결과에 반영되지 않았다. 따라서 높은 스코어-낮은 스코어를 가진 쌍에서 대부분의 관계를 찾아낼 수 있었으며, 그 결과 Faster R-CNN만을 사용한 접근법과 비교하여 7.3%의 성능 향상을 이룰 수 있었다.

V. 결론

본 논문에서는 객체와 객체의 관계를 활용하여 이미지에 존재하는 객체의 인식 성능을 향상시키는 방법을 소개하였다. 제안 방법에서는 Faster R-CNN을 이용한 객체 인식과 VTransE를 활용한 관계 학습 방법을 결합한 후 모바일 앱 UI의 위젯 인식에 맞는 제약조건을 가하여 UI를 구성하는 위젯을 인식하고자 하였다. 제안 방법을 사용하지 않았을 때는 위젯 인식률이 86% 정도로 자동 테스트에 적용하여 실제 앱을 테스트하기에는 부족한 성능을 보였지만 제안 방법을 적용한 결과 여전히 자동 테스트에는 부족하나 인식률이 93%로 향상됨을 확인하였다. 추후 연구에서는 기존의 제약조건보다 유효한 제약조건을 추가하여 위젯 인식률을 현장 활용이 가능한 수치로 향상시키는 방법을 연구할 것이다.

References

- [1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)*, pp.580-587, 2014.
- [2] R. Girshick, "Fast R-CNN," in *Proc. of the International Conference on Computer Vision (ICCV 2015)*, pp.1440-1448, 2014.
- [3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.39, no.6, pp.1137-1149, 2016. DOI:10.1109/TPAMI.2016.2577031

- [4] J. Johnson, R. Krishna, M. Stark, L.-J. Li, D. A. Shamma, M. S. Bernstein, and L. Fei-Fei, "Image retrieval using scene graphs," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, pp.3668-3678, 2015.
- [5] C. Galleguillos, A. Rabinovich, and S. Belongie, "Object categorization using co-occurrence, location and appearance," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, pp.1-8. 2014.
DOI:10.1109/CVPR.2008.4587799
- [6] W. Choi, Y.-W. Chao, C. Pantofaru, and S. Savarese, "Understanding indoor scenes using 3D geometric phrases," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013)*, pp.33-40, 2013.
- [7] C. Lu, R. Krishna, M. Bernstein, and L. Fei-Fei, "Visual relationship detection with language priors," in *Proc. of the European Conference on Computer Vision (ECCV 2016)*, pp.852-869, 2016.
- [8] G. Gkioxari, R. Girshick, and J. Malik, "Contextual action recognition with R*CNN," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, pp.1080-1088, 2015.
- [9] V. Ramanathan, C. Li, J. Deng, W. Han, Z. Li, K. Gu, Y. Song, S. Bengio, C. Rossenber, and L. Fei-Fei, "Learning semantic relationships for better action retrieval in images," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, pp.1100-1109, 2015.
- [10] B. Dai, Y. Zhang, and D. Lin, "Detecting visual relationships with deep relational networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, pp.3076-3086, 2017.
- [11] L.-C. Chen, A. G. Schwing, A. L. Yuille, and R. Urtasun, "Learning deep structured models," in *Proc. of the International Conference on Machine Learning (ICML 2015)*, pp.1785-1794, 2015.
- [12] H. Wang, X. Shi, and D.-Y. Yeung, "Relational deep learning: a deep latent variable model for link prediction," in *Proc. of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, pp.2688-2694, 2017.
- [13] M. Long, Z. Cao, J. Wang, and P. S. Yu,

"Learning multiple tasks with multilinear relationship networks," in *Proc. of the Thirty-first Annual Conference on Neural Information Processing (NIPS 2017)*, pp.1593-1602, 2017.

[14] J.R.R. Uijlings, K.D.A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *INT J COMPUT VISION*, vol.104, no.2, pp.154-171, 2013.
DOI:10.1007/s11263-013-0620-5

[15] K. Mao, M. Harman, and Y. Jia, "Sapienz: multi-objective automated testing for Android applications," in *Proc. of 2016 International Symposium on Software Testing and Analysis*, pp.94-105, 2016.

[16] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. of the Twenty-seventh Conference on Neural Information Processing Systems (NIPS 2013)*, pp.2787-2795, 2013.

[17] H. Zhang, Z. Kyaw, S.-F. Chang, and T.-S. Chua, "Visual translation embedding network for visual relational detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, pp.5532-5540, 2017.

BIOGRAPHY

Min-Su Park (Member)



2017 : BS degree in Computer and Communication Engineering, Kangwon National University.

Ho-Sik Seok (Member)



1999 : BS degree in Computer Engineering, Seoul National University.
2001 : MS degree in Electrical Engineering and Computer Science, Seoul National University.
2012 : PhD degree in Electrical Engineering and Computer Science, Seoul National University.
2001~2004 : Instructor, Korea Military Academy.
2016~present : Assistant professor, Dept. of Computer and Communications Engineering, Kangwon National University.