

## SEMINAR ANNOUNCEMENT

### DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Faculty of Engineering

Website: <https://www.eng.nus.edu.sg/ece/>

**Area: Signal Analysis & Machine Intelligence**

**Host: Dr Feng Jiashi**

<b>TOPIC</b>	:	<b>Visual Relationship Detection with External Visual-Linguistic Knowledge from Multimodal Representation</b>
<b>SPEAKER</b>	:	<b>Mr. Chiou Meng-Jiun Graduate student, Computer Science, NUS</b>
<b>DATE</b>	:	<b>Friday, 17 July 2020</b>
<b>TIME</b>	:	<b>4.50PM to 5.30PM</b>
<b>WEBINAR</b>	:	<b>Join Zoom Meeting <a href="https://us02web.zoom.us/j/497447113?pwd=b1V4ZFJGaEZuTTJjZUVlekJCRkhpQT09">https://us02web.zoom.us/j/497447113?pwd=b1V4ZFJGaEZuTTJjZUVlekJCRkhpQT09</a> Meeting ID: 497 447 113 Password: 027731</b>

### ABSTRACT

Over the past few years, increasing attention has been paid to visual relationship detection, which aims to reason over relationships among salient objects in images. Inspired by human reasoning mechanism, we believe commonsense knowledge is essential for reasoning visual relationships of objects in images, which is however rarely considered in existing methods. In this paper, we propose a transfer learning based approach named Relational Visual-Linguistic Bidirectional Encoder Representations from Transformers (RVL-BERT), which performs relational reasoning with both external visual and language commonsense knowledge based on pre-trained multimodal visual-linguistic representations. In particular, RVL-BERT utilizes an effective spatial module and a novel mask attention module to better capture spatial information among the objects based on external knowledge. Moreover, it decouples object detection from visual relationship recognition by taking in object names directly so that it can be used on top of any object detection system. We show through quantitative and qualitative experiments that, with the novel spatial and mask attention modules, RVL-BERT surpasses previous state-of-the-art on two challenging visual relationship detection datasets.

### BIOGRAPHY

Meng-Jiun Chiou is a computer science PhD student at National University of Singapore and is under the supervision of both Prof. Roger Zimmermann at the Department of Computer Science and Prof. Jiashi Feng at the Department of Electrical and Computer Engineering. His research is focused in computer vision and machine learning, especially, scene understanding with visual relationship detection. He also has experience in visual relationship guided image captioning, zero-shot indoor localization with graph neural networks and weakly labeled audio tagging.

<https://www.eng.nus.edu.sg/ece/highlights/events/>