

# Causality Inspired Retrieval of Human-object Interactions from Video\*

Liting Zhou<sup>†§</sup>, Jianquan Liu<sup>‡¶</sup>, Shoji Nishimura<sup>‡</sup>, Joseph Antony<sup>†</sup>, Cathal Gurrin<sup>†</sup>

<sup>†</sup>Dublin City University, Ireland

<sup>‡</sup>Biometrics Research Laboratories, NEC Corporation, Japan

<sup>§</sup>Email: [liting.zhou@insight-centre.org](mailto:liting.zhou@insight-centre.org)

<sup>¶</sup>Email: [j-liu@ct.jp.nec.com](mailto:j-liu@ct.jp.nec.com)

**Abstract**—Notwithstanding recent advances in machine vision, video activity recognition from multiple cameras still remains a challenging task as many real-world interactions cannot be automatically recognised for many reasons, such as partial occlusion or coverage black-spots. In this paper we propose a new technique that infers the unseen relationship between two individuals captured by different cameras and use it to retrieve relevant video clips if there is a likely interaction between the two individuals. We introduce a human object interaction (HOI) model integrating the causal relationship between the humans and the objects. For this we first extract the key frames and generate the labels or annotations using the state-of-the-art image captioning models. Next, we extract SVO (subject, verb, object) triples and encode the descriptions into a vector form for HOI inference using the Stanford CoreNLP parser. In order to calculate the HOI co-existence and the possible causality score we use transfer entropy. From our experimentation, we found that integrating casual relations into the content indexing process and using transfer entropy to calculate the causality score leads to improvement in retrieval performance.

## 1. Introduction

With the increasing number of surveillance cameras around us, there is a growing need for searching and inferring the context of the captured events and the potential relationship that may exist between the humans, objects, and their interactions. Despite many advancements in computer vision such as improved instance and human action recognition, surveillance video understanding and prediction still remains a challenging task. In order to bridge the gap between the semantic understanding and vision representation, it is necessary to understand the causality [6] between different human and object interactions.

Causal reasoning is the process of identifying causality: the relationship between a cause and its effect. Humans can reason about many topics with the aid of causal understanding and this naturally translates into semantic activity-based queries in retrieval systems. For instance, consider a security professional who is interested in finding surveillance video that shows a stolen item (e.g. handbag) being passed from one individual to another, or in other video retrieval applications (e.g. sports) where a human activity is a topic of

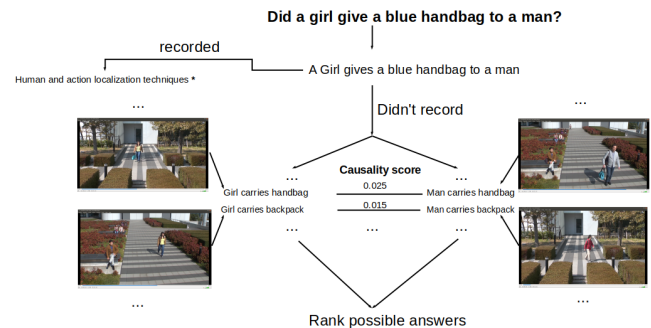


Figure 1. An example of the challenge that shows a stolen item (e.g. handbag) being passed from one individual to another: two individuals carrying the same blue backpack, with a probable handover of the backpack not captured by the surveillance cameras.

interest, or even in personal data analytic (e.g. lifelogging) when a person wishes to query past activities captured using wearable cameras.

Learning causality has been applied in myriad of high-impact applications that include computer vision [17, 23], financial data analysis [24, 25], environmental health [21, 26] and education [22, 27]. Also, causality has been viewed as a significant component to improve the accuracy of information prediction and retrieval [28, 29]. Many approaches in recent times combine machine learning with causal inference [3] to answer causal questions [17] and to build a causal inference framework to answer questions regarding the causal relationships between object statuses and human actions, which in many cases may be hidden or not visible. In this paper, we report the performance of a first-generation approach to solve this challenge, demonstrating that the proposed model can successfully infer causality between the different human-object interactions in video data. Figure 1 summarises the research challenge visually, which shows two individuals carrying the same blue backpack, with a probable handover of the backpack not captured by the surveillance cameras.

Extending prior research, we propose a human-centric approach to causality inference which finds the potential human-object interactions in videos. The proposed model contains a number of components, but has been designed to

be flexible and component agnostic. The main contributions of this paper can be summarised thus:

- the integration of causal inference into predictive modeling and producing a causality human-object interaction model.
- a calculation of the probable causality between different human-object interactions.
- a method for qualifying which causal relations were found in a video and matching causal relations between those expressed in videos and in user queries.
- identification of the causal structure between different human object interactions and the application in the information retrieval domain.

## 2. Related Work

### Human Object Interaction

#### Image Captioning with human and object detection

Automatic image and video captioning is important, but it is challenging, because it connects two major artificial intelligence fields: computer vision and natural language processing. The state of the art in this [29] uses recurrent neural networks for this, which is based on long short-term memory units as decoder. We choose this model to extract sentences followed by the extraction of the HOI triplets.

**Causality Inference in Computer Vision** The concept of “Causality” has been applied in different research domains, for instance: identification of causality relations between events [13, 14], commonsense causal reasoning in texts [15, 16]. In the last decade, there has been preliminary research applying causality in computer vision for a deeper video context understanding, perceptual causality are learned from video [17] through a framework of causal structure between “actions” and “fluents” in videos. In contrast to these works, our goal is to perform causal analysis in general classes of video content with the object and action extracted via deep learning framework.

**Visual Relationship Detection** Visual relationship detection is an instance of computer vision research and currently the main focus is to describe the interaction between pairs of objects that are located in the same images [19]. There are not many reported approaches in the literature for finding and inferring these visual relationship between the humans and objects from multi-camera videos.

## 3. Application Scenario

In order to find a solution to the problem of causality detection, we present a short scenario in the domain of security/surveillance video, with a typical layout of security cameras. In an outdoor environment, there is a U-shaped pathway which is configured with 3 cameras: 1,2 and 3, as shown in Figure 2. As can be seen, some corners (top left) are not covered by the cameras, and are in effect hidden regions, which we refer to as ‘dead-zones’. While cameras 1, 2 and 3 are separately recording their own field-of-view, they will not capture activities in the dead-zones. Consider

a scenario in which there is a person (A) carrying a blue handbag, which is given to a person (B) in a dead-zone (i.e. not captured by any surveillance camera). Camera 3 captures footage of a person carrying a similar blue handbag (the same bag) that was carried by person A, a few moments previously. Without causality detection, this handover is not identified by the security system, which could potentially pose a major security risk. An information need such as “*did a person give a blue handbag to another?*” can not be answered without a human-object interaction causality model. Our challenge, in this work, is to find and infer causality between unseen actions, based the video captured from two or more different cameras.

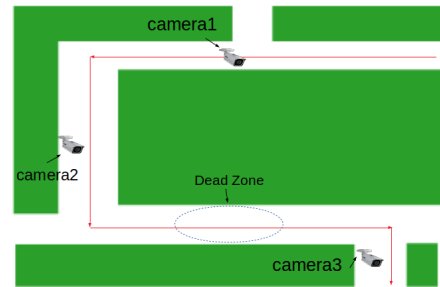


Figure 2. Scenario Map showing ‘dead zones’.

In order to tackle the problem highlighted by the above scenario, we propose a HOI (Human-Object Interaction) causality model to infer unseen interactions. The overview of the HOI causality module is shown in the Figure 3, which we use as the basis for the following explanation of the model. The framework is a pipeline consisting of four components: keyframe extraction SVO extraction, encoder & HOI model, HOI inference. Note that in this proposed approach, we select state-of-the-art components, but further research could be done to enhance the performance of each component, thereby enhancing the performance of the overall technique. The detailed descriptions are listed in section 4.

## 4. Human-object Interaction Causality Model for Retrieval

The proposed HOI causality model operates by identifying human actions and contextual features from video frames. The process can be described as follows: firstly extract the instance-level appearance features such as object colour, shape, human face, and gender using a standard process to distinguish humans and objects. Secondly, with a goal to dynamically generate a causality map conditioned on the different HOIs corresponding video frames are retrieved from the dataset based on a generated causality matrix, which is a graph encoding interactions between known objects in the video. The detailed explanation of techniques are explained in the following sections. We provide an overview of the entire approach in section 4.1 to section 4.4 and outline the method for the inference in section 4.5.

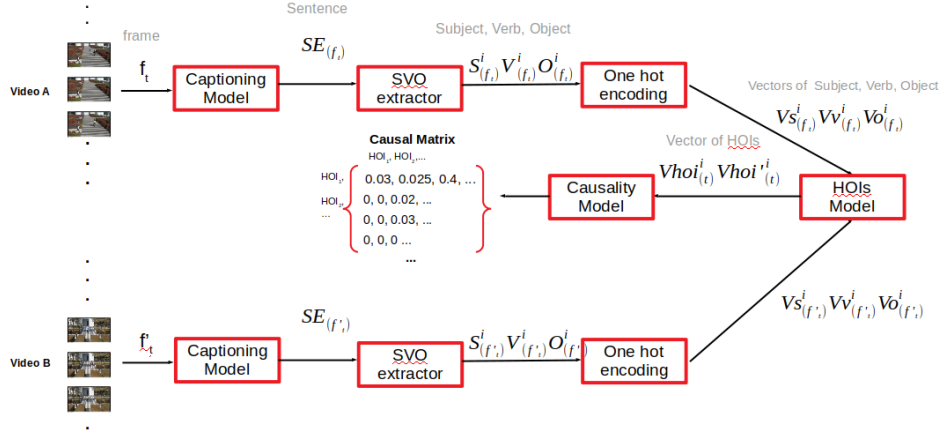


Figure 3. Overview of the proposed model.

#### 4.1. Keyframe Extraction

The first step is the extraction of keyframes:  $f$  from the source videos as shown in Figure 3. Keyframe extraction is a commonly-used tool that assists the organisation of video content by selecting a set of summary keyframes (one keyframe per 0.5s) to represent video sequences in a storyboard form. After this step, we use a captioning model to extract the human object interactions from these keyframes, which can be viewed as a multiple-point process for qualifying the causal relationship between them.

#### 4.2. HOI extraction using Image Captioning Model and ICAN model

The second phase of the process (see Figure 3) is image captioning, which generates labels or annotations for the previously extracted keyframes  $SE_f$ . We consider the problem of HOI generation as being similar to the problem of video captioning. Currently, the state-of-the-art video captioning results are based on using LSTMs and CNNs [31]. In our implementation, we use CNNs to extract image features and a multiple LSTM structure to extract the main image caption features and employ multi-layer LSTMs with 512 dimensional hidden states to model the probability of different words in the captions. The reason to choose this captioning model is that we want to use a multi-point process to qualify the existence of HOIs and apply this multi-point process in causality transfer theory. Hence, using the image captioning models is a pre-processing step in our method.

The inputs to the HOI generation process are two surveillance videos captured using different cameras that are located in different locations, but within close proximity to one another, and with a dead-zone between their fields-of-view. The output of the captioning model is a vector of sentences describing the keyframes in terms of objects and activities (along with their scores) from each video sequence. Like the scenario described earlier in which two

people are seen to be holding the same blue bag in different camera sequences. The challenge to be addressed now is how to infer the reasons why the bag has been transferred from one person to another; was the bag handed over, or was it potentially stolen? This corresponds to the SVO extraction process that is explained in the next section. An example of the image caption for a keyframe based on image captioning model is also given in next section.

To investigate if there is improvement in the performance, another tool is also implemented to extract Human-object interaction using H-iCAN model which is designed based on iCAN model [12] and an instance-centric attention network for detecting and recognising the interaction between each person and the object they are interacting with. An example of the HOI generated for a keyframe (using the iCAN model described later in the paper) is shown in Figure 4.



Figure 4. HOI output example for one keyframe showing objects and activities with scores

#### 4.3. SVO Extraction

SVO stands for subject, verb, object triplets ( $\langle S, V, O \rangle$ ) that are extracted from the sentences. After generating the descriptions for videos in the previous HOI generation process, we analyze these descriptions via HOI reasoning. The first step is to extract the triplets that will be used to calculate the causality from these descriptions. We choose

to use the state-of-the-art stanford natural language parser created by Stanford CoreNLP [18] for this task, which generates an appropriate grammatical structure of sentences, for instance, which groups of words go together (as “phrases”) and which words are the subject or object of a verb. Given an input image as shown in Figure 4, the output of this SVO Extraction process would be a sentence in the form “a man is walking on the sidewalk with a blue bag in hand”. The next phase (shown on the right side of Figure 3) is to encode the textual descriptions from SVO extraction into a vector form for HOI inference).

#### 4.4. Encoder & HOI model

The input for the causality model must be a sequence of time-series data, hence we have to encode the  $\langle S, V, O \rangle$  triplets into vectors which can be used to represent the presence of a Subject, Action, and Object triplet. The chosen encoder method is one-hot encoding, which is one type of vector representation of text for machine learning. The one-hot encoding is a representation of categorical variables as binary vectors and each word value is represented as binary vector that encodes zero values for all items except the index of the chosen word.

After extracting the binary vectors from the SVO, we employ the HOI model to combine the presence of Subject, Action, and Object triplets. The combination of the same HOI at nearby times extracted from different videos, is the main novelty of our HOI model. Given video A captured by camera 1 from time  $t_s$  to  $t_e$ , and video B captured by camera 2  $t_s$  to  $t_e$ , the challenge is how to combine these into HOI models. The detail of how to process the combination is concatenating the vectors from the two videos sequences during the same time period and adding them together when the subjects, actions or objects are same.

#### 4.5. HOI Inference

HOI inference part is concerned with how to quantify relationship between different HOI triplets which are in the form:  $\langle Human, Action, Object \rangle$ , extracted from the image captions. In order to calculate the level of relationship between different HOI values, we utilise a method named transfer entropy [4] to count the co-existence of Subject, Object, Actions and then compute the possible causality score between them. The causality of the time dimension imposes that past actions have consequences for subsequent actions or events. While the human and object appearance features contain strong cues for recognising the interaction, using appearance features alone often leads to possible inference results being missed.

**4.5.1. Non-parametric Transfer Entropy.** Wiener’s causality theorem [32] states that one variable could be regarded as the cause of another if the ability to predict the future of the second variable is enhanced by utilising information concerning the first variable. Hence, we consider to apply transfer entropy to infer the relationship

between different HOIs and we create the causality score matrix (labeled CSM in formula 1) to show the causality between them. We use transfer entropy and obtain a scalar causal score from each causal measure by integrating over the frequency of occurrence. As we described in section 4.4, the decoder we choose is one-hot encoding, which uses binary labels to denote the existence of the human Subject, Actions and Objects. After processed by the decoder, the vector will be transferred into the HOI model, which adds the vectors to the other HOI vectors extracted from different video to produce the relatively complete HOI vector. For example:  $I = \langle x_1, x_2, \dots, x_n, \dots \rangle$ ,  $J = \langle y_1, y_2, \dots, y_n, \dots \rangle$ . The causality model will calculate the transfer entropy score using formula 2 and create the causality matrix.

$$S_{TE} = \operatorname{argmax}(CSM(HOI_i, HOI_j)) \quad (1)$$

The detailed *transfer entropy* is defined as:

$$TE_{HOI_j \rightarrow HOI_i} = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) \log \left( \frac{p(x_{n+1} | x_n, y_n)}{p(x_{n+1} | x_n)} \right) \quad (2)$$

### 5. Experimental Results

We evaluate the performance of our proposed HOI model in comparison to a state-of-the-art approach (Huaman-Centric iCAN [11] model), using both our HOI-S dataset containing short video clips and public datasets from youtube (HOI-Y).

#### 5.1. Experimental dataset

Firstly, for training both the iCAN model and image captioning models we used the following two public datasets: **V-COCO**[9] is a dataset that provides detailed HOI annotations. It includes a total of 10,346 images containing 16,199 human instances. Each person is annotated with a binary label vector for 26 different actions. Each person can perform multiple actions at the same time, for example a man is walking and holding a handbag.

**HICO-DET** is a subset of the HICO dataset [7]. It contains 600 HOI categories over 80 object categories and provides more than 150K annotated instances of human-object pairs.

And for evaluating the HOI inference retrieval, we used the following two datasets:

**HOI-S dataset:** this is a dataset that we generated for this work, which was collected by volunteers in Japan using three different surveillance cameras arranged in a formation similar to that shown in Figure 2. It includes 3.9G of video data and we annotated these dataset using the LabelImg annotation tool. The dataset shows an instance of a man holding a bag who swaps his bag with the bag of another. In order to evaluate our approaches, we trimmed this video and cut them into 2 test sequences.

**HOI-Y dataset:** this is a dataset we downloaded from YouTube and we trimmed the video data to the HOI dataset as per requirements. It contains one 6 minute and one 2 minute videos labelled by the LabelImg annotation tool, which we used to test the entire procedure of our method.

## 5.2. Implementation details

We evaluate three methods based on our work for the experiments of this proposed approach:

- The first method is using the image captioning model which is trained on GPU Quadro P5000 using CNN and LSTM, as described in this paper (referred to as **LSTM+CNN** in the following result tables);
- The second method called **H-iCAN** is based on iCAN [11] with a feature backbone of ResNet 50 FPN. The H-iCAN is a modified iCAN that better fits our learning and prediction framework. The main purpose of the modification is generating the HOI output directly from video frames and filtering all HOIs to choose the high-ordered human-centric ones which will be used to represent the frames. We trained the upgraded H-iCAN model using the v-coco and HICO-DET datasets and applied the trained model to extract the HOI triplets which are used to calculate the causality. Then, we extract the HOI of HOI-S and HOI-Y datasets to retrieve the relevant human object interactions;
- The final method is a **Human-in-Loop**, defined as a model with human interaction, which is employed in order to integrate a best-possible HOI detection, to check the captioning result and give relevant feedback which will be used to modify the caption results.

## 5.3. Results Comparison

In this section, we compared our proposed framework and evaluated our approach based on the retrieval results from the three approaches previously described. As discussed in Section 4, the ultimate goal of our model is to infer information transfer between different HOIs in the videos dataset by analysing the relationship between objects and humans. Specifically, we compared the approaches (described in previous section) using the HOI-S and HOI-Y datasets in terms of F1 measure and Rank@10.

The sample of transfer entropy scores calculated based on HOI inference results obtained from the H-iCAN model is shown in the Table 1 and the corresponding illustrations of the extracted images are shown in Figure 5. Additionally, the evaluation results of retrieval performance on the HOI-S dataset is shown in Table 2 for all three approaches described. As can be seen and expected, the Human-in-Loop approach significantly outperforms the two automated approaches, which highlights the importance of highly effective HOI generation process. Additionally, the H-iCAN approach clearly outperforms (0.21 compared to 0.14) the LSTM-CNN approach. Table 3 highlights the retrieval performance of both automated approaches on the HOI-Y dataset, which shows a significantly closer performance between the two approaches. The reason for this is being explored, but our conjecture is that the difference in performance is due to the nature of the videos in the HOI-Y dataset not being as suitable for the HOI extraction process. Finally Table 4 shows direct comparison between the two automated approaches in terms of F1 measure across both datasets and gives a clear indication of how much

TABLE 1. RESULTS USING HUMAN ANNOTATED HOI-S DATA

| HOI_start               | HOI_end                   | TE_score |
|-------------------------|---------------------------|----------|
| Woman 1_carries_handbag | Man 4_carries_backpack    | 0.063    |
| Woman 1_carries_handbag | Man 1_carries_backpack    | 0.060    |
| Woman 1_carries_handbag | Man 5_carries_backpack    | 0.041    |
| Woman 1_carries_handbag | Woman 2_carries_backpack  | 0.041    |
| Woman 1_carries_handbag | Man 6_carries_handbag     | 0.024    |
| Woman 1_carries_handbag | Man 6_carries_shoulderbag | 0.018    |

TABLE 2. COMPARISON RESULTS IN HOI-S DATASET

| Approach      | Rank@10    |
|---------------|------------|
| LSTM+CNN      | 0.14       |
| H-iCAN        | 0.21       |
| Human-in-Loop | <b>0.8</b> |

the H-iCAN approach performs better than the LSTM-CNN approach.

## 6. Discussion & Conclusion

In this paper, we propose a human-centric causality model for HOI retrieval. Our core idea is to retrieve semantically meaningful results from different but relevant camera videos. Our work highlights the appearance of a person and an object instance, which allow us to gather relevant contextual information facilitating HOI detection. Our experimental results show that our two automated approaches, the LSTM+CNN and H-iCAN are significantly outperformed by an approach based on the Human-in-Loop annotation of the HOIs prior to retrieval. This is certainly what one would expect and gives a motivation to further enhance our HOI generation process. The two automated approaches are compared and the H-iCAN approach performs better than LSTM+CNN, but this in itself is not surprising because the quality of the HOI annotations from the H-iCAN approach is significantly better than those generated by the LSTM+CNN.

We view this work as an initial implementation and proposal for a method for inferring the relationship (causality) between humans and objects in video data. The three approaches evaluated highlight the scope for significant enhancement to retrieval performance for this important and challenging topic. Additionally, this work has focused on a class-agnostic instance-centric approach. We believe that the class-dependent instance-centric attention is a promising future direction, which we will explore, along with the implementation of an end-to-end video retrieval system for various use-cases, such as surveillance video, conventional video retrieval, and for the semantic retrieval of lifelog data. Finally we intend to integrate transfer entropy calculation into the H-iCAN model to further enhance the retrieval performance by reducing the noise inherent in the process by reducing the complexity of the approach.

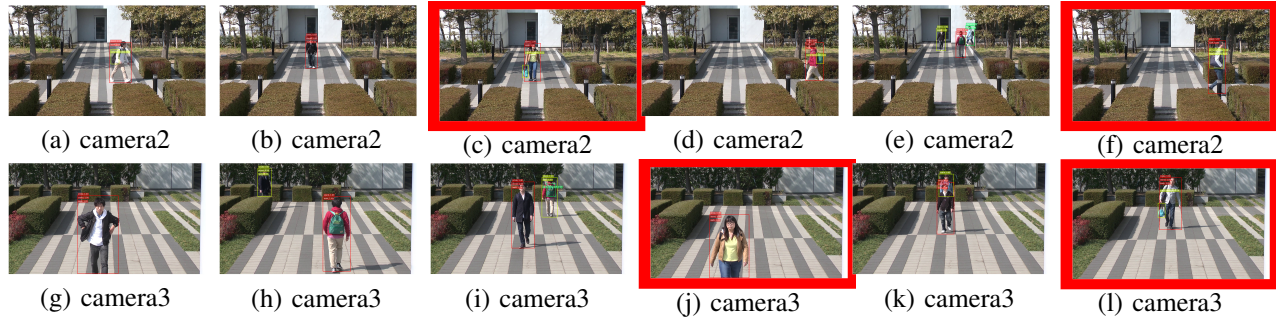


Figure 5. Result used the H-ICAN model and HOI-S dataset

TABLE 3. COMPARISON RESULTS IN HOI-Y DATASET

| Approach | Rank@10 |
|----------|---------|
| LSTM+CNN | 0.14    |
| H-iCAN   | 0.15    |

TABLE 4. COMPARISON BETWEEN BOTH AUTOMATED METHODS ON BOTH DATASETS

| Approach | F1   |
|----------|------|
| LSTM+CNN | 0.13 |
| H-iCAN   | 0.25 |

## Acknowledgments

This publication has emanated from the research at NEC Corporation reserving all rights of the contents, and also was partially supported by research grants from Irish Research Council (IRC) under Grant Number GOIPG/2016/741.

## References

- [1] Lu, Cewu, et al. "Visual relationship detection with language priors." European Conference on Computer Vision. Springer, Cham, 2016.
- [2] Hu, Zhichao, and Marilyn A. Walker. "Inferring Narrative Causality between Event Pairs in Films." arXiv preprint arXiv:1708.09496 (2017).
- [3] Prabhakar, Karthir, et al. "Temporal causality for the analysis of visual events." 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010.
- [4] Schreiber, Thomas. "Measuring information transfer." Physical review letters 85.2 (2000): 461.
- [5] Ver Steeg, Greg, and Aram Galstyan. "Information transfer in social media." ACM WWW, 2012.
- [6] Guo, Ruocheng, et al. "A Survey of Learning Causality with Data: Problems and Methods." arXiv preprint arXiv:1809.09337 (2018).
- [7] Chao, Yu-Wei, et al. "Hico: A benchmark for recognizing human-object interactions in images." Proceedings of the IEEE International Conference on Computer Vision, 2015.
- [8] Yang, Zhongliang, et al. "Image captioning with object detection and localization." International Conference on Image and Graphics. Springer, Cham, 2017.
- [9] Gupta, Saurabh, and Jitendra Malik. "Visual semantic role labeling." arXiv preprint arXiv:1505.04474 (2015).
- [10] Xu, Bingjie, et al. "Interact as You Intend: Intention-Driven Human-Object Interaction Detection." arXiv preprint arXiv:1808.09796 (2018).
- [11] Gao, Chen, Yuliang Zou, and Jia-Bin Huang. "ican: Instance-centric attention network for human-object interaction detection." arXiv preprint arXiv:1808.10437 (2018).
- [12] Chao, Yu-Wei, et al. "Learning to detect human-object interactions." 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.
- [13] Hu, Zhichao, and Marilyn A. Walker. "Inferring Narrative Causality between Event Pairs in Films." arXiv preprint arXiv:1708.09496 (2017).
- [14] Hashimoto, Chikara, et al. "Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features." Computational Linguistics, Vol. 1. 2014.
- [15] Luo, Zhiyi, et al. "Commonsense causal reasoning between short texts." Principles of Knowledge Representation and Reasoning, 2016.
- [16] Yeo, Jinyoung, et al. "Machine-Translated Knowledge Transfer for Commonsense Causal Reasoning." AAAI Artificial Intelligence, 2018.
- [17] Fire, Amy, and Song-Chun Zhu. "Learning perceptual causality from video." ACM TIST 7.2 (2016): 23.
- [18] Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. 2014. Computational Linguistics: System Demonstrations, pp. 55-60.
- [19] Zhu, Yaohui, and Shuqiang Jiang. "Deep structured learning for visual relationship detection." AAAI Artificial Intelligence, 2018.
- [20] Pearl, Judea. "Theoretical impediments to machine learning with seven sparks from the causal revolution." arXiv preprint arXiv:1801.04016 (2018).
- [21] Li, Jundong, Osmar R. Zaane, and Alvaro Osornio-Vargas. "Discovering statistically significant co-location rules in datasets with extended spatial objects." Data Warehousing and Knowledge Discovery. Springer, Cham, 2014.
- [22] Meitriana, Made Ary, I. Kadek Rai Suwena, and Luh Indrayani. "The Influence of Fraud Triangle and Theory of Planned Behavior on Students Academic Fraud in Bali." TEAMS 2018. Atlantis Press, 2019.
- [23] Yang, Chao-Han Huck, et al. "When Causal Intervention Meets Image Masking and Adversarial Perturbation for Deep Neural Networks." arXiv preprint arXiv:1902.03380 (2019).
- [24] Sun, Yunchuan, Yufeng Shi, and Zhengjun Zhang. "Finance Big Data: Management, Analysis, and Applications." (2019): 9-11.
- [25] Bdowska-Sjka, Barbara, and Agata Kliber. "The causality between liquidity and volatility in the Polish stock market." Finance Research Letters (2019).
- [26] Bind, Marie-Able. "Causal modeling in environmental health." Annual review of public health 40 (2019).
- [27] Fetters, Michael D., and Jos F. Molina-Azorin. "In This Issue: Innovations in Mixed Methods Causality, Case Study Research With a Circular Joint Display, Social Media, Grounded Theory, and Phenomenology." (2019): 123-126.
- [28] Lei, Lei, et al. "Event Prediction Based on Causality Reasoning." Asian Conference on Intelligent Information and Database Systems. Springer, Cham, 2019.
- [29] Kayesh, Humayun, Md Islam, and Junhu Wang. "On Event Causality Detection in Tweets." arXiv preprint arXiv:1901.03526 (2019).
- [30] Kang, Le, et al. "Construction of fast retrieval model of e-commerce supply chain information system based on Bayesian network." Information Systems and e-Business Management (2019): 1-18.
- [31] Xu, Kaisheng, Hanli Wang, and Pengjie Tang. "Image captioning with deep LSTM based on sequential residual." IEEE ICME(2017).
- [32] Bressler, Steven L., and Anil K. Seth. "WienerGranger causality: a well established methodology." Neuroimage 58.2 (2011): 323-329.