Real-time Semantic Healthcare System: Visual Risks Identification for Elders and Children

Malak Belkebir[1], Toufik Messaoud Maarouk[2], Brahim Nini[1]
[1]Research Laboratory on Computer Science's Complex Systems ReLa(CS)2, University of Oum el Bouaghi, Oum el Bouaghi, Algeria
[2]ICOSI laboratory, Dept. of Mathematics and Computer Science, Khenchela University, Khenchela, Algeria
E-mail: belkebir.malak@univ-oeb.dz, maarouk.toufik@univ-khenchela.dz, brm.nini@gmail.com

Keywords: deep learning, ontology, healthcare system, risk identification, high-level semantic, reasoning

Received: May 26, 2024

Deep learning and data-driven approaches are commonly used to avoid accidents involving elders and children. However, existing models are limited by a semantic gap, hindering their ability to infer new risks that have not been previously trained. In this paper, a real-time healthcare system is developed to identify and infer visual risks in surveillance videos for elders and children. The system consists of three main modules: "visual information extraction," which leverages advancements in artificial vision techniques such as GCD and IoU for relationship detection, YOLO for object detection, and ResNet18 for scene recognition; "ontology modeling," where a new high-level ontology named "Risks-Identification-Onto" is constructed based on FOL and DLs; and "risk identification," where the system infers risks by deducing new knowledge through the reasoning of generated formal rules over the data-driven techniques. Additionally, the system generates a high-level semantic description of the risky situation. Four common risk scenarios - "Hurt," "Burn," "Existing-in-dangerous-places," and "Hit"- are selected to evaluate the effectiveness of the proposed system. Evaluation is conducted using the Charades and A2D datasets, each including 9,848 and 3,782 indoor activity videos. It demonstrate the system's efficiency in identifying and inferring risks in real-time with an accuracy ranging from 97.61% to 99.43% for each scenario.

Povzetek: Študija predstavlja sistem zdravstvenega varstva v realnem času za prepoznavanje vizualnih tveganj pri starejših in otrocih, ki uporablja globoko učenje in ontologijo.

1 Introduction

Parents today are more distracted by work, outside activities, and other responsibilities, making it difficult to provide constant care for their young children or aging parents. Many statistics [1, 2] indicate an increase in incidents involving elders and children, underscoring the importance of ongoing supervision. Children, in particular, are more prone to accidents due to their inherent curiosity and lack of knowledge about potential risks. On the other hand, elderly people with dementia or mobility issues have a higher risk of having an accident. With the prevalence of indoor and outdoor surveillance cameras, such as those found in homes, offices, and public places, they are widely used for manually monitoring children and elders, despite the fact that it is time-consuming and potentially ineffective during periods of inattention.

One solution to automate the monitoring process is through artificial vision based on data, also known as "data-driven approaches," which have demonstrated significant breakthroughs and results [3–5]. These methods sought to limit risks by tracking individuals, identifying objects, and recognizing actions/behaviors. However, their reliance solely on existing data and the need for extensive training and testing datasets renders them inadequate for inferring and identifying emerging risks in line with evolving safety and care standards for children and the elderly. The primary limitations of data-driven approaches lie in the "semantic gap" and the absence of knowledge-based reasoning regarding new information. For example, while a data-driven approach can detect "an elder near a table, a knife is on the table," it lacks the ability to infer from this output that "this individual may be at risk of harm from the sharp object (knife)." Moreover, following a training phase with a large dataset, a single model or algorithm is typically employed to detect a specific risk in a given setting.

Another solution used by researchers [3, 4, 6–14] is the integration of data-driven approaches with ontology and logical techniques to minimize the resource-intensive requirements, such as data and computational power. An ontology, as defined by Borst [15], is a "formal specification of a shared conceptualization," providing rich meanings and semantics for a specific domain that computers can understand and use in formal ways. This integration effectively reduces the semantic gap and has been applied across various fields, such as image retrieval, object recognition, and risk prediction. However, it is still uncommon for real-time detection of dangers affecting children and the

elderly.

Moreover, these studies do not consider the deduction of new risks in real-time, which is the most effective preventative measure for unexpected accidents. For that, this paper introduces a real-time healthcare semantic system that combines formal approaches with artificial vision techniques for the identification and inference of visual risks in surveillance videos for elders and children. The proposed system uses the least amount of resources and data possible; thus, its performance in real-time is effective and appropriate for the sake of risk identification. Additionally, it presents a newly constructed ontology called Risks-Identification-Ontology. On the one hand, data-driven approaches are used to extract visual data such as objects (YoLoV5), visual relationships (Grounding Consistency Distillation GCD), spatial-geometric relationships (IoU), and scene environment (ResNet18). The outputs are represented as sets of triples. On the other hand, a combination of formal approaches, including ontology, FOL, and description logics (DLs), is employed for knowledge representation, along with reasoning-logic rules (i.e., those generated with a high level of semantics) to detect and infer dangers.

The real-time risk identification process for each scenario is accomplished with no need for a training phase, and this is done by mapping the set of triples obtained to the developed ontology, which serves as a ”Fact Base.” If the situation is deemed dangerous, each person in the scene will be assigned to the appropriate risk class by applying a reasoner to the well-established rules using the Semantic Web Rule Language SWRL, which serves as the ”Rule base.” Additionally, an auto-description is generated, along with an alert in case of danger.

The contributions of this paper are: 1) the construction of a new ontology called ”Risks-Identification-Onto,” and 2) the development of a high-level semantic healthcare system combining logic with artificial vision, while using minimal resources and maintaining real-time performance, as well as closing the semantic gap between information -i.e., results of data-driven approaches- and knowledge -i.e., results of reasoning about information-. The effectiveness of the proposal is demonstrated by its use in this critical domain, assisting parents and caregivers in ensuring the safety of elders and children with an accuracy of 97.61% to 99.43 %. Notably, four common risk scenarios are identified: ”hurt,” ”burn,” ”existing in dangerous environments,” and ”hit.”

The rest of this paper is structured as follows: Section 2 summarizes the state-of-the-art work. Section 3 describes the proposal’s architecture and methodology, together with the newly developed ontology. Section 4 illustrates the study cases, experiments, and tests. Finally, Section 5 includes a conclusion.

2 Related works

Several studies on risk identification and people-care are being conducted, with various approaches being proposed. These approaches can be categorized into three main groups: imaging and artificial vision approaches, i.e., data-driven-based, formal approaches, i.e., knowledge-driven based, and hybrid approaches, i.e., combining both. In the following, a synthesis of each approach is outlined and addressed.

2.1 Data-driven based for people-care and risk identification

Data-driven approaches for risk identification have demonstrated high accuracy in various fields, including monitoring systems [3] that provide valuable assistance to parents in monitoring infants to prevent accidents and unforeseen injuries. The author in [4] introduced an improved accident prediction model that combines the temporal pyramid of the LSTM (TP-LSTM) model, the temporal attention mechanism, and the early exponential loss (EEL) function to anticipate infant accidents within seconds or fractions of a second before they occur. Similarly, the work in [3] addresses a monitoring system where risk detection considers the spatial interactions between each newborn and adjacent objects, such as entering dangerous zones or coming into contact with harmful objects that should be predefined in each time and case. The proposed system in [16] introduces a wearable device equipped with a fall detection approach. This device takes the form of a wireless bracelet and it is designed to aid individuals with vision impairments by detecting obstacles in indoor environments. In contrast, [17] describes ”Friendy,” a deep learning-based chatbot. This chatbot is intended to provide psychotherapy interventions to children with autism. Finally, [18] proposes a monitoring model that tracks pedestrian flow to prevent crowding and stampedes.

Finally, in [19], a deep learning-based system is proposed for monitoring shared autonomous vehicles. It employs three distinct algorithms: a system for detecting violent actions, a system for detecting violent objects, and a system for detecting lost items.

Despite the significant results of these works, they continue to present obstacles and challenges in terms of semantic reasoning and the inference of new information or knowledge that differs from the inputted training data.

2.2 Knowledge-driven based for people-care and risk identification

Ontologies are becoming increasingly popular in knowledge-driven approaches. For instance, the study in [7] models two ontologies of actions and objects for elders in the home setting. Its purpose is to formally describe the scope domains while providing additional semantic details about them. [6] proposes an ontology for representing

and identifying risks during building renovations.

Nonetheless, relying solely on an ontology for risk identification provides semantic modeling of a specific domain without the auto-extraction of real-world information. Consequently, it may be unable to automatically determine hazards in real-world situations without the manual input of data.

2.3 Hybrid approaches based for people-care and risk identification

Although ”data-driven approaches” achieved significant results, they are semantically low/medium-level, lacking the ability to reason and infer new high-level semantic knowledge and interpretations. State-of-the-art works recommend integrating ”knowledge-driven approaches” with them. The authors of [11] have synthesized literature supporting this integration and have demonstrated how formal and logical inferences can bridge the semantic gap.

Furthermore, many works proposed aim to extract semantic visual relationships in sports images [8], semantic analysis for human behavior [14], enhancing image recognition [13], and risk prediction [12]. In addition, a multimodal approach is proposed in [10] with the aim of identifying hazards at building sites. Finally, [9] proposed a graph-based framework that integrates linguistic Natural Language Processing (NLP), OpenPose, and YoLov4, with a reasoning approach to process regulatory rule sentences and images for on-site occupational hazards like ”working on height” and ”operating a grinder.”

2.4 Recap

The majority of the discussed works, which combine ”knowledge-driven” and ”data-driven” approaches, have effectively bridged the semantic gap. However, their applications in real-time danger recognition for children and elders remain limited (Table 1). To address this gap, this work proposes a semantic system integrating artificial vision techniques with knowledge-driven methods, which will be discussed in more detail in the following sections.

3 Architecture and methodology of the proposal

This work combines deductive reasoning from knowledge-driven approaches with inductive reasoning from data-driven methods. The proposed real-time semantic system integrates artificial vision with logic and ontology to detect dangers affecting children and elders in indoor/outdoor environments by integrating low/medium-level semantic information with high-level knowledge. The architecture consists of three main modules, as shown in Figure.1:

1. Visual Information Extraction: This module identifies visual elements within a captured scene, including the indoor/outdoor environment, individuals, surrounding objects, and their visual relationships;

2. Ontology Modeling: A formal ontology named ”Risks-Identification-Onto” is developed to interpret the results of the first module. It renders them machine-readable for formal interpretations and prepares them for semantic-risk reasoning and querying. Additionally, the outputs of the visual information extraction module are exploited to instantiate individuals in the ontology;

3. Risk Identification:First-order logic (FOL) and Description logic (DLs) are used to define common risk scenarios, which serve as inputs for the risk inference process, along with the ontology and its instantiation outputs. Risky scenarios posing threats to children and/or elders are inferred by applying reasoning-logic rules to the extracted visual information.

3.1 Visual information extraction

This module aims to extract the minimum amount of information essential for effective risk detection while maintaining real-time aspects. This information is used to generate low/medium-level semantic description of a particular scene in indoor/outdoor environments. The following are the models used to detect visual content: deep learning for object detection (YoLov5), scene recognition (Resnet18), and visual relationships detection (GCD), with IoU metric for spatial-geometric extraction. The outputs are represented in sets of three triples, a computational format, and then stored in JSON files for use as inputs in subsequent modules.

3.1.1 Objects detection and scene recognition

The YOLO algorithm, a recent convolutional neural network (CNN) model, excels at speed and precision for object recognition and is widely used in action recognition, and risk analysis. Since its inception by [20], YOLO has evolved through versions like YOLO V2 to V8 [21].

The authors of YOLO revamped object identification from classification to regression, replacing a two-stage algorithm with one-stage methods. Unlike earlier methods, which required hundreds or thousands of passes per image, YOLO conducts detection in a single pass by dividing the image into grid regions. Each region predicts bounding boxes (BBOX) and probabilities, indicating object classes, locations, and scores. The YOLO network consists of three main phases [22, 23]:

- Backbone: a convolutional neural network that extracts features from various sizes of images;

- Neck: series of network layers aggregate the extracted image features to enrich semantic information, serving as input to the prediction layer;

Table 1: Recap of related works

Ref.	Proposed	Methods	Dataset	Type, time, and accuracy-rate of inference (respectively)	Limitation
[3]	Monitoring system for detecting accidents involving infants in rooms	OpenPose for individual detection, background subtraction technique	The article does not specify a dataset, but it does use an infant doll measuring 58cm tall	Induction (no reasoning about knowledge). Unspecified Time and Accuracy	Preliminary experiments with no knowledge inference
[4]	Early accident prediction model for infants and children	TP-LSTM, exponential loss (EEL) function, TWO-STREAM-CONVNET	Baby Video Dataset (BVD)	Induction. 4.196 seconds. 61.13%	The model prediction is limited to trained cases
[16]	Fall detection wireless bracelet for vision-impaired individuals. It is based on the detection of obstacles in indoor environments	Firebase database, NodeMCU WiFi, HC-SR04 ultrasonic distance sensor	Not specified	Induction (for training of obstacles). 0.3 seconds. Accuracy is Not mentioned (demonstrates real-world experiments)	Several devices are used for the detection of environmental objects, which should be wearable constantly, with no inference on risks
[17]	”Friendy”, A therapy enhancement framework for autistic children. It is based on deep learning with a contextual chatbot	LSTM, the Gated Recurrent Unit (GRU) topology	Newly constructed dataset	Induction (based on information provided by experts). Mentioned as real-time but not calculated. Accuracy of 80.5%	The lack of data makes the system challenging to scale and integrate into diverse real-world settings
[19]	Violence monitoring system for shared autonomous vehicles	YOLOv5 for object detection, 3D ConvNet, SlowFast, and Temporal Segment/ Shift Networks for video action recognition	TAO, COCO, MoLa InCar	Induction. Real-time (170-330). Accuracy of 94.32%	Limited to the trained samples
[7]	Construction of two formal ontologies of home actions and objects for elders.	Standard ontology construction	Charades, Home-Ontology	Deduction (logical inference). Unspecified Time. Accuracy cannot be estimated; logic-based results are always true	The mere use of ontology without evaluation in real-world scenarios is insufficient

[6]	Risk ontology for building renovations	Standard ontology construction	Deep renovation projects data (RINNO; Europe 2020 research project)	Deduction. Unspecified Time. Accuracy cannot be estimated	Effectiveness has not been demonstrated in real-world projects
[8]	Semantic extraction and interpretation of visual relationships in sports images	Standard ontology construction, VGG-16 for object detection, VRD for relationship detection	HCVRD	Induction for object detections, deduction false positive filtering. Unspecified Time. Accuracy Depends on each ”Concept”	It lacks the inference of new knowledge
[14]	Chatbot for personality disorder assistance through semantic analysis	NLP, Standard ontology construction	Twitter	Inductive, deductive. Unspecified Time. Accuracy of 72%	High potential for misinterpretation of disorder intricacies
[13]	Image recognition enhancement with ConSE -a new ontology- of digital images in construction sites	CNN-LSTM, Graph Neural Network (GNN), standard ontology construction	Construction site images produced by authors	Induction for ontology development and deduction for ontology validation. Unspecified Time and Accuracy	Manual low-level information determination in the images, and lack of system evaluation
[12]	Disease prediction model	LSTM, Bidirectional Gated Recurrent Unit (Bi-GRU) for the prediction	Not mentioned	Induction for features extraction and training, deduction for diseases prediction. Unspecified Time and Accuracy	Training such models effectively requires large and diverse datasets
[10]	Hazards identification at building sites; multimodal approach	Standard ontology construction, Swi-prolog, Nlp	VRD	Induction (information detection), deduction (logical semantic reasoning with swi-prolog). Unspecified Time. Accuracy of 49.91%	The majority of the computational capacities and resources are concentrated on safe objects, which limits real-time risk identification
[9]	Graph-based hazard identification framework for occupational sites that integrates linguistic and visual information	OpenPose, YOLOv4, spaCy (feature generation), NetworkX (graph structure)	Created by authors and presents ”working on height” and ”operating a grinder”	Induction (detection of visual and linguistic information), deduction (hazard reasoning). Unspecified Time and Accuracy	High computational requirements; not real-time results

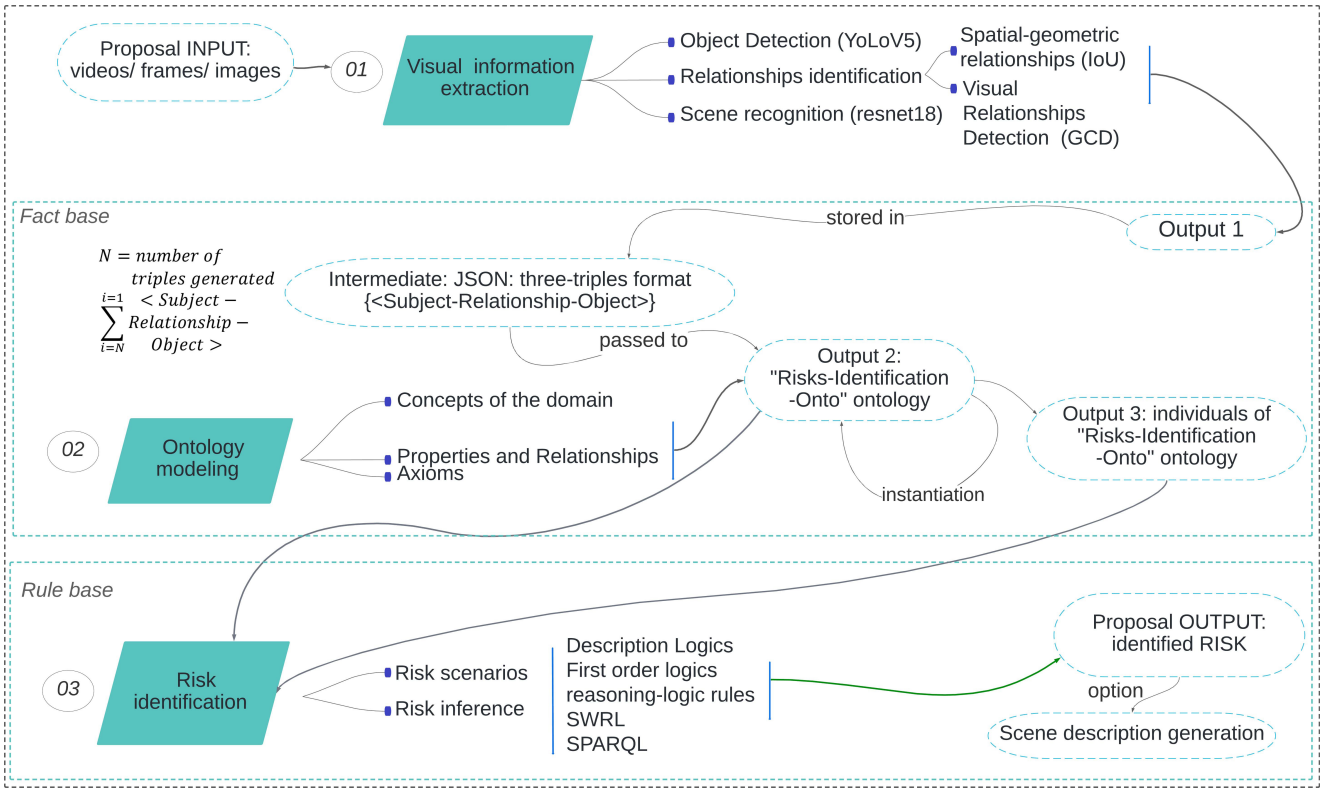

Figure 1: Architecture of the proposed approach.

– Head: predicts inputs from the neck, providing object classes with coordinates, probabilities, and scores.

In this work, YOLOv5 [24] is used for object detection in indoor/outdoor environments. It is pretrained on the COCO dataset and implemented using the PyTorch framework. YOLOv5 incorporates the AutoAnchor algorithm by Ultralytics, which adjusts anchor boxes to achieve a better fit for the database and the training parameters. The architecture incorporates a modified CSPDarknet53 backbone, a stem, and a convolutional layer with a large window size to save memory and compute during feature extraction. A spatial pyramid pooling fast (SPPF) layer accelerates computation by combining features into a fixed-size map. Each convolution goes through batch normalization (BN) and SiLU activation. While the neck makes use of SPPF and a modified CSP-PAN. This model outperforms existing classifier-based techniques. Its advantage lies in its exceptional speed, enabling real-time results with high precision, making it particularly suitable for risk identification proposals.

On the other hand, the localization of the detected objects is recognized using the pretrained model (ResNet18) on the PLACE 365 [25] dataset for scene recognition. This model accurately identifies the environment in which a person is situated within a scene.

3.1.2 Relationships identification

Two types of relationships have been identified to recognize interactions between each pair (person, objects): (1) spatial-geometric relationships and (2) the visual relationships detection (GCD) model.

Spatial-geometric relationships

The Intersection over Union (IoU) metric is the evaluation standard for quantifying the degree or ratio of overlap between two BBOXs [26], which means that it operates directly and instantaneously on the BBOXs. The IoU can be calculated as shown in Equation.1 by dividing the intersection of the two bounding boxes by their union. The metric used in this work considers three possible ratios: (0), [0-1], and (1). These ratios represent three types of spatial-geometric interactions between individuals and scene objects: ”far,” ”overlap,” and ”complete overlap,” respectively (see Figure.2).

$$IoU(x,y) = \frac{Intersection\ of\ the\ two\ BBOXes\ \|X \cap Y\|}{Union\ of\ the\ two\ BBOXes\ |X \cup Y|} \quad (1)$$

Visual Relationships Detection (VRD)

Several approaches and models have demonstrated remarkable outcomes in visual relationship detection (VRD) [27]. These approaches reveal the visual interactions between each pair of detected objects. Formally, in an image IMG with $Objects$ representing the number of detected ob-

Figure 2: Results applying IoU; the Bounding Box of "person1" is overlapping the Bounding Box of "motorcycle2."

jects, and P denoting the total number of all potentially created pairs of $Objects$ (referred to as $Pair$), the relationship is defined in Equation.2:

$$P = Objects \times (Objects - 1), \quad Pair = \langle Subject - Object \rangle \tag{2}$$

The objective of VRD is to generate $< Subject - Predicate - Object >$ triples and/or scene graphs, with the predicate representing the relationship between the subject and object (e.g., $< Person - next_to - knife >$). This work adopts the Grounding Consistency Distillation (GCD) model [28] to identify visual relationships between persons and nearby objects, particularly those that may pose a danger. To streamline the process and maintain real-time performance, only triples involving individuals and potentially hazardous items are selected, with benign objects filtered out during the information extraction phase.

The GCD model, a semi-supervised distillation training approach, addresses a key weakness in traditional Scene Graph Generator (SGG) and VRD models. The SGG models often prioritize high recall over the expense of considering spatial and visual evidence, relying heavily on datasets biased toward common relationships, termed "bias on relationships." Consequently, they may generate inaccurate predictions by disregarding visual information, spatial coordinates, and genuine object connections (see Figure.3.a). For example, if the dataset used for training consists of numerous instances of a person carrying a knife, it will be deemed that every person can hold a knife. As a result, if a knife is detected alongside multiple individuals, , the model may incorrectly identify the entire group as wielding the knife, regardless of their original proximity to the object. This can result in false-positive risk identification alarms, which are exacerbated by a lack of "negative examples" to highlight those false facts, as "not everyone in the scene necessarily is holding the detected knife."

This is a positive reason to confirm that the GCD is a viable model for this proposal. It can accurately identify which object is in-relation-to the subject based on its position, geometric coordinates, and visual information, as well as, having high accuracy and recall outcomes (see Figure.3.b). To achieve this purpose, three networks are used: the Grounder, the teacher-SGG, and the student-SGG. More specifically, using a pretrained grounding network, the teacher-SGG is constrained to predicting the most pertinent and ground relationships to the scene to create spatial common sense knowledge, which is subsequently distilled into the student-SGG model. The latter considers unlabeled data and provides out-of-distribution cases that cast doubt on the perception of the network of the dominant classes. Additionally, there is a generation phase of negative labels for the unlabeled data (see Figure.3.c).

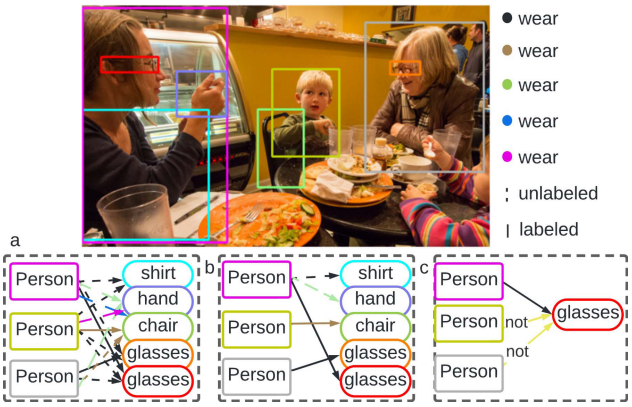

Figure 3: a. Training results using standard "bias on relationships" datasets, demonstrating "overfitting."b. The training results of GCD demonstrate that this issue has been addressed. c. excerpt from the generation of negative labels.

3.1.3 Representation format of visual information (triples)

The nature of extracted visual information, being heterogeneous-textual, poses challenges for integration with ontology for real-time formal manipulation and reasoning. To enable computer comprehension, a structured and unified format is necessary. A three-triples structure ($< Subject - Relationship - Object >$) is adopted to encapsulate this information, which is then outputted in JSON format. This JSON file serves as an intermediary representation and a mapper between the modules of the proposed system. Figure.4 showcases a sample output from the "visual information extraction" module.

3.2 Ontology modeling

The challenges of data-driven approaches include constructing logical systems and generating high-level semantic knowledge in real-time with high precision. One solution to these challenges is to integrate ontology, First Order Logic (FOL), and Description Logics (DLs). Researchers [11] confirm that this is effective for describing and providing formal reasoning for semantic content.

{ "FrameContext": {
"nameC": "kitchen1"
},
"Spatial-geometric_relationships": [
["person1","far","oven1"],
["person1","overlap","diningtable5"]
["person1","far","knife2"],
["person1","overlap","oven3"],
["person1","overlap","microwave4"]],
"GCD_relationships": [
["stove3","next to","person1"],
["knife2","on","diningtable5"], }

Figure 4: Results of visual information extraction module.

Due to that, a new ontology named "Risks-Identification-Onto" is developed and integrated with the Visual Information Extraction module. The idea is to minimize the semantic gap, provide a rich context and deep meaning for its findings, and automatically generate high-level semantic descriptions of the given scene. The construction of this ontology is grounded in the foundational principles articulated by Gruber i.e., "explicit specification of a conceptualization" and Borst's "formal specification of a shared conceptualization" [29].

From a formal description perspective, the present ontology is constructed with a concise approach: conceptualizing aspects of interest and their intra/inter relationships, formalizing these concepts using appropriate language, and achieving a "shared conceptualization" stage in which primitives are understandable to ontology users. Four background knowledge are chosen: (1) the subsumption of classes, defining the connection where one class is a subclass of another; (2) the domain/range restrictions, which specify the domain or range of object classes for a relation class; (3) the cardinality restrictions, limiting the maximum number of relations of a certain relation class that an object may have; and (4) object collections, referring to groups of image objects that fall under the same object class.

From a logical implementation perspective, this ontology uses DLs and FOL to generate a machine-readable structure for a particular domain. It abstractly -explicitly or implicitly- conceptualizes all Concepts (Cp) with their Properties and the Relationships (R) between them. Axioms (ϕ), which impose constraints on these entities, are also integrated, along with Individuals serving as instances (I) (Equation.3). This formalization aims to unify domain knowledge, derive new knowledge through logical inference, and facilitate automated reasoning and querying processes.

$$O = / \left\{ \begin{array}{c} \Sigma\phi \\ Cp = \{cp1, cp2, \ldots, cpn\} \\ R = \{r1, r2, \ldots, rn\} \\ I = \{i1, i2, \ldots, in\} \end{array} \right\} \quad (3)$$

The "Risks-Identification-Onto" is constructed through the following steps:

- A: Define the domain of ontology.
- B: Search for existing ontologies to reuse.
- C: Select the taxonomy of the chosen domain.
- D: Define the top-concepts, then categorize step (C) into "concepts" and "relationships, i.e., properties."
- E: Instantiate individuals based on the results of visual information extraction module, using intermediate JSON files as an input.

Since no relevant ontology exists for risk identification in indoor/outdoor environments, the top-down" approach [30] is used for the construction of the "Risks-Identification-Onto." It entails starting with the top-level concepts and gradually refining them to establish a hierarchical structure. These concepts are conceptualized based on the definition of the formal extensional and intentional concepts: $A = (D, R, C, \mathfrak{S}, \mathfrak{A})$, with:

$$\begin{array}{r} \forall d_i.\top \sqcap d_j.\top \vDash \\ D \sqcap D(d_i) \vDash \Xi \sqcap D(d_j.\top) \vDash \Xi \sqcap d_i.\top \equiv \neg d_j.\top \end{array} \quad (4)$$

$$\exists r_{i,j} \sqcap r_{j,i} \vDash R \sqcap R(r_{i,j}) \sqcap R(r_{j,i}) \vDash \Xi \sqcap r_{i,j} \equiv \neg r_{j,i} \quad (5)$$

$$\begin{array}{r} \forall r_{i,j} \sqcap r_{j,i} \vDash R \sqcap R(r_{i,j}) \sqcap R(r_{j,i}) \vDash \Xi \sqcap d_i.\top \sqcap \overrightarrow{r_{i,j}} d_j.\top \\ \equiv \neg d_j.\top \overrightarrow{r_{j,i}} d_i.\top \end{array} \quad (6)$$

Noting that:

(1) D represents the set of defined aspects/concepts. The YOLO and VRD BBOXES, with PLACE-365 labels, are chosen as the ontology's taxonomy and will play the roles of concepts representing $< Subject >$ *and* $< Object >$. Figure.5 depicts ten classes of the top layer: 1) Thing, 2) Be_alive, 3) Environment, 4) Food, 5) Furniture, 6) Mean_of_transport, 7) Object_to_use, 8) Positioning, 9) Traffic_lights and 10) is_in_Danger. Figure.6 shows their properties, while Figure.7 illustrates additional classes derived from the top layer, among which, elder, child, adult, Animal, Plant, Safe_Outdoor, Unsafe_Outdoor, Safe_Indoor, Unsafe_Indoor, ground_transportation, Maritime_ transportation, Air_ transportation, Sport_equipment, General_things, Electric_device, Electromechanical_device, Electronic_device, Kitchen_tool, Sharp_tool, Hot_tool, being_in_dangerous_place, Burn, Hurt, and Hit.

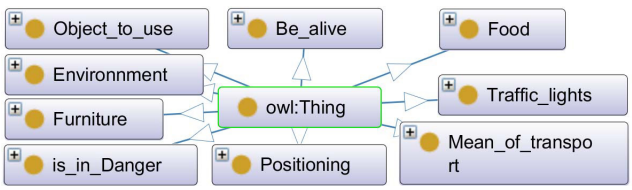

Figure 5: The top-layer concepts of the Risks-Identification-Onto; OntoGraph.

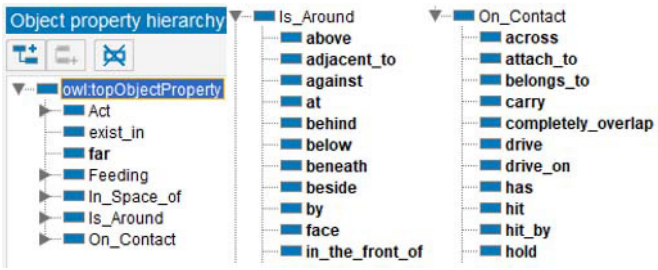

Figure 6: An excerpt from the object properties classification of the Risks-Identification-Onto.

Additionally, according to the National Institutes of Health [31], "age" is defined with a "restriction and reasoning on numbers" - a Python module that includes functions for managing numerical constraints and performing reasoning tasks with numerical data- to automatically classify a person as "elder," "adult," or "child," as shown in Table 2.

Table 2: The "restriction and reasoning on numbers" applied to automatically classify a person to "elder," "adult" or "child"

```
class Elder(Person): equivalent_to=
[Person & age.some
(ConstrainedDatatype
(int,min_inclusive = 65))]
class Child(Person): equivalent_to=
[Person & age.some
(ConstrainedDatatype
(int, max_inclusive = 12))]
class Adult(Person): equivalent_to=
[Person & age.some
(ConstrainedDatatype
(int,min_inclusive =13,
max_inclusive = 64))]
```

(2) C is the constraint between D and R. For instance, taking Figure.4, let the concepts used for axioms generation be: $d1$, $d2$, and $d_3 \vDash D$, and $rd1,d2,rd1,d3,rd3,d1$,and $rd3,d2 \vDash R$. Where $d1$=($Person1$, …, $Person_n$), and $d2$=($oven1$, …, $oven_n$), $d3$ = ($knife1$,…, $knife_n$). Relationships that only exist between these concepts, i.e., person, knife, and oven, are $rd1,d2$=($hold$, far, $overlap$, $next\_to$, on, …), $rd1,d3$=($near$, on, $next$,…), $rd3,d1$=(far, $overlap$, $next\_to$, …), $rd2,d1$=far, $overlap$, $next\_to$, on, …), $rd3,d2$= ($next\_to$, on, far, $overlap$, …), $rd1,d2$, $rd1,d3$, $rd3,d1$, and $rd3,d2$.

In addition, the conceptualization should remain unchangeable with changes in world instantiation [32]. The verb "hold" serves as an example; the axioms and rules that define the verb "hold" should not change with changes in the environment (and vice versa; "hold" is understood with the same axioms and rules, for example, a "knife" cannot "hold" a "person"). The Risks-Identification-Onto is recorded under these restrictions, which limit and provide extensive background knowledge for all and between aspects and relationships.

(3) $\mathfrak{S}$ is the ontology universe defined as $\mathfrak{S} = \{\xi^{'},\xi^{''},\xi^{'''},\dots\}$. This means that all the ontology entities are built in accordance with the time evolution for each existence of the world ontology, and it is defined based on the following formal description:

$$\xi^{'} \rightsquigarrow t_{,} \vDash T \sqcap \xi^{''} \rightsquigarrow t_{,,} \vDash T, if \xi^{'} \equiv \neg\xi^{''}$$
$$\exists R.\top \sqsubseteq A \sqcap \exists d_i.\top \sqcap d_j.\top \vDash D \sqcap \exists r_{i,j} \vDash R \wedge \xi^{'}$$
$$\xi^{'}(d_i.\top \overrightarrow{ri,} \overrightarrow{jd_j}.\top) \equiv \neg\xi^{''}(d_i.\top \overrightarrow{ri,} \overrightarrow{jd_j}.\top)$$

(4) $\mathfrak{A}$ restricts and defines the conceptualization between the aspect sets D and the ontology universe $\mathfrak{S}$. It is preferable to consider the unary conceptualization of aspects and the binary intra/inter-relationships as more rigid to build a straightforward formal extensional of aspects. e.g., $Person1, Person2, oven1, oven2, knife1, knife2$. Additionally for $overlap, far, hold, next_{to}, on, near$. It is constructed and mapped to the same extensions as the ontology universe for this reason. Similar assumptions were used to build the formal intentional of aspects:

$$\exists\xi^{'} \sqcap \xi^{''} \sqcap \xi^{'''} \sqcap \dots \vDash \mathfrak{S} : Person1(\xi^{'}) \equiv d1 \wedge Person1(\xi^{'''}) \equiv d1 \wedge Person1(\dots) \equiv d_1$$
$$\exists\xi^{'} \sqcap \xi^{''} \sqcap \xi^{'''} \sqcap \dots \vDash \mathfrak{S} : oven1(\xi^{'}) \equiv d_2 \wedge oven1(\xi^{'''}) \equiv d_2 \wedge oven1(\dots) \equiv d_2$$
$$\exists\xi^{'} \sqcap \xi^{''} \sqcap \xi^{'''} \sqcap \dots \vDash \mathfrak{S} : knife1(\xi^{'}) \equiv d_3 \wedge knife1(\xi^{'''}) \equiv d_3 \wedge knife1(\dots) \equiv d_3$$

$$\exists overlap(\mathfrak{u}_{d1,d2}) \vDash \mathfrak{A} \sqcap (\xi^{'} \sqcap \xi^{''} \sqcap \dots) \vDash \mathfrak{S} \equiv \{(Person1(\xi^{'}) \sqcap oven1(\xi^{'})) \wedge (Person1(\xi^{''}) \sqcap oven1(\xi^{''})) \wedge (Person1(\dots) \sqcap oven1(\dots))$$

$$\exists hold(\mathfrak{u}_{d1,d3}) \vDash \mathfrak{A} \sqcap (\xi^{'} \sqcap \xi^{''} \sqcap \dots) \vDash \mathfrak{S} \equiv \{(Person1(\xi^{'}) \sqcap knife1(\xi^{'})) \wedge (Person1(\xi^{''}) \sqcap knife1(\xi^{''})) \wedge (Person1(\dots) \sqcap knife1(\dots))$$

The Risks-Identification-Onto includes 504 classes, 88 properties, 1307 axioms, and 710 logical axioms. Top-layers of the Risks-Identification-Onto are shown in Figure.7.

After completing steps A to D, step E involves automatically instantiating individuals and relationships for each captured scene. This process applies uniformity to the results obtained from the "visual information extraction" module. The instantiation is depicted in Figure.8, and it serves as input for the "risk inference" module. Consequently, the proposal can deduce risks using sets of triples

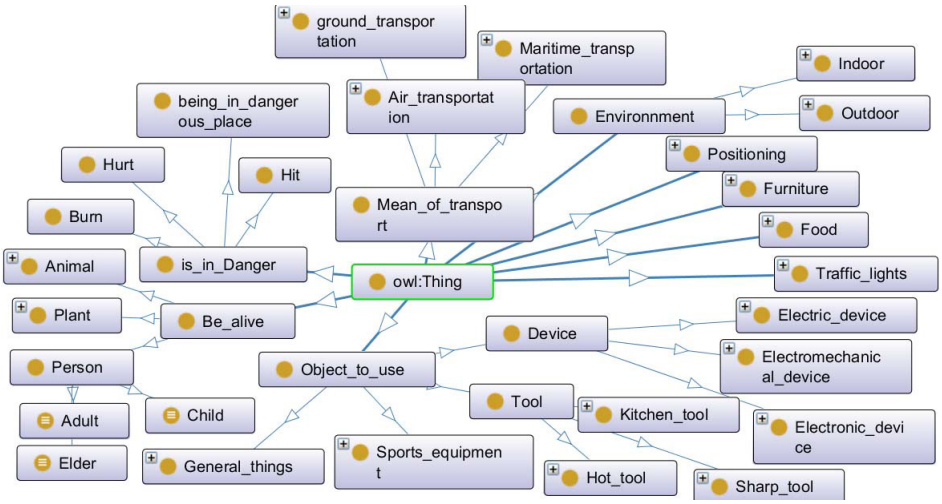

Figure 7: Part of the Risks-Identification-Onto concepts hierarchy; OntoGraph.

$< Subject - Relationship - Object >$ without requiring a separate training phase for each scenario.

Moreover, the inference outcomes are used to auto-generate descriptions and trigger alarms, providing users with semantically descriptive information to facilitate quick understanding and intervention.

3.3 Risks identification

The module consists of two primary steps: defining risk scenarios and performing risk inference. In the initial step, potential risk situations are outlined and described using FOL-based DLs, which are well suited to be integrated with the ontology to leverage its logical inference and reasoning capabilities. In the subsequent step, the system automatically deduces and identifies situations that may endanger elders and children in both indoor and outdoor environments through logical reasoning.

DLs are formal languages that focus on knowledge representation, inference, and reasoning. It employs FOL to formalize and describe Knowledge Bases (KB), which include, in this case, concepts C, relationships R, individuals I and axioms ϕ [33]. KB contains three types of entities [34]:

1. Constants: set of individuals $\{c1, c2, \ldots cn\}$ e.g., "$person1$," "$knife1$."

2. Unary relations: set of concepts $\{cp1, cp2, \ldots, cpn\}$, e.g., "$Person$," "$knife$."

3. Binary relations: roles and properties, e.g., age, $overlap$, $In\_Contact$.

DLs is composed of the two groups of axioms (denoted ϕ), which is the Fact Base specifying entities of a given knowledge domain with their constraints: $KB =< A, T >$ [34]:

1. Assertional axioms A: named $ABox$, sets of individuals I assertions, e.g., "Person(person1), age(person1, 70)"

2. Terminological axioms T: named $TBox$, complex descriptions of relationships R between concepts Cp and collections of inclusion assertions, e.g., $Elder \sqsubseteq Person, Elder \equiv Person \sqcap age \geqslant 65$.

In this work, the $ABox$ and the $TBox$ are generated as follows:

1. The set of triples $< Subject - Relationship - Object >$ is mapped to binary relations $Relationship(Subject, Object)$ according to D, $R \models \Xi$, and $\mathfrak{A}$ in $\mathfrak{S}$.

2. Constants $< Subject >$ and $< Object >$ are asserted to their parent concepts using the unary relation $Cp(C)$ and/or inclusion assertions, according to C in terms of $\mathfrak{A}$ and the time evolution of $\mathfrak{S}$.

For instance, "$knife1$" is instantiated as a C of the Cp "knife," denoted by the axiom $knife_C\ (knife1)_{Cp}$. It is considered both a $Sharp\_tool$ and a $Kitchen\_tool$, symbolized by $knife \sqsubseteq Sharp\_tool, knife \sqsubseteq Kitchen\_tool$.

Rules of danger, referred to as the Rule Base, are defined and generated by integrating and formalizing the knowledge of the $ABox$ and the $TBox$, as depicted in Figure.9.

3.3.1 Risk scenarios

The paper outlines four primary risk scenarios, highlighting the most common potential dangers faced by "elderly" or "child" individuals (we hypothesized that adults can protect themselves in normal circumstances):

– Hurt: When $Person(P1) \geqslant 65$- or $Person(P1) \leqslant 12$-, i.e., $Elder$ or a $Child$. They are susceptible to injuries from sharp tools like $knife$ or $scissor$ under

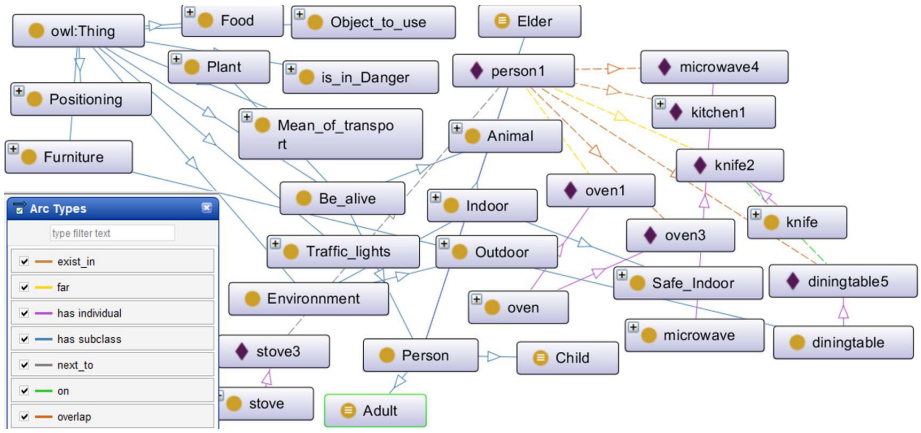

Figure 8: Results of the instantiation of Risks-Identification-Onto ontology of the frame in Figure.4; OntoGraph.

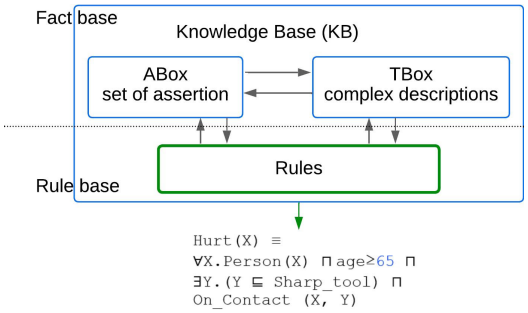

Figure 9: Sample of a defined Rule using FOL.

the following conditions: 1) Direct contact with sharp tools, indicated by spatial relationships such as $R \models \Xi$: "*overlap*," or "*completelyoverlap*," or via a VRD relationship, indicating a *Contact* relationship between $Person(P1)$ and a *Sharp_tool*. 2) Indirect contact R with *Sharp_tools*, such as when they are placed on furniture that individuals come into contact with.

- Burn: *Elders* and *Children* may suffer burn from *Hot_tools* such as microwave or oven in such situations: 1) Direct contact with hot tools, as indicated by spatial relationships like $R \models \Xi$: "*overlap*," or "*completelyoverlap*" or any "*Contact*" relationship with a *Hot_tools*. 2) Proximity to hot tools, as indicated by the "*Is_Around*" relationship.

- Hit: Individuals may be struck by $ground\_transportation \sqsubseteq means\_of transport$ under the following circumstances: 1) Detected *Elder* and *Child* who have the relationship $R \models \Xi : on$ the *street* and have a *contact* relationship with any $means\_of transport$, but are not inside or on it. 2) When an *Elder* has dementia, which requires close supervision, is riding a $motorcycle, bicycle$, or *bike*.

- Existing in dangerous places: *Elder* or a *Child* should not be present in hazardous outdoor or indoor environments such as $cliff, physicslaboratory, etc.$, as these are only suitable for adults and experts.

3.3.2 Risk inference

This module defines the logical-based rules that use the outputs of data-driven approaches as input for deductive reasoning. It provides high-level semantic understanding and reasoning capabilities that closely resemble human deduction in the field of real-time risk identification.

The "Rule Base" that formalizes the risk scenarios is generated using the Semantic Web Rule Language (SWRL). SWRL is both an extension of OWL and a rule language. It can be used in conjunction with ontology to automatically express more sorts axioms and constraints due to its strong deductive inference and reasoning abilities [35].

The SWRL rules are expressed in high-level abstraction as *OWL* Concepts Cp, properties/Relationships R, and instances/individuals I. Each rule consists of two parts: a consequent part called the Head, which is a set of atomic formulas that can serve as the logical conclusion of reasoning, and an antecedent part called the Body, which is a conjunction of atomic formulas. Equation.7 presents a standard SWRL rule ($->$ is a separator between the Head and the Body):

$$A(?i1) \wedge B(?i1, ?i2) \rightarrow C(?i1) \tag{7}$$

A and Cp are OWL classes; $A(i1)$ and $Cp(i1)$ are atoms; B is a property; $i1$ and $i2$ are OWL individuals; and $?i1$ and $?i2$ are SWRL variables. In contrast, $A(?i1) \wedge B(?i1, ?i2)$ is only valid and true if both $A(?i1)$ and $B(?i1, ?i2)$ are true. In this case, when the logical conclusion of the inference over the body is reached, the fact base is expanded to include the newly inferred and deduced one. For example, the scenario "Hurt," presented with FOL (Equation.8), can be generated as Rule 2:

$$FOL: Hurt(X) \equiv \forall X.Person(X) \sqcap age \geqslant 65 \sqcap \exists Y.(Y \sqsubseteq Sharp\_tool) \sqcap On\_Contact(X,Y) \quad (8)$$

$$SWRL: Rule_1: Person(?x) \wedge age(?x,?b) greaterThan(?b,64) \wedge Sharp\_tool(?c) \wedge On\_Contact(?x,?c) \rightarrow Hurt(?x)$$

If a given person (x) in the given scene is more than or equal to 65 years old, i.e., an elder, and a detected sharp tool(c) comes into contact with person (x), then, (x) is in danger of being hurt and will be reclassified as *Hurt* class; the *Hurt* class is expanded to include this type of individual.

Following that, the proposal uses the Protocol And RDF Query Language SPARQL [36] to enable automatic responses to the query "Is anyone in danger?." Table 3 exemplifies the query, "Is there anyone who could be hurt?"

Table 3: The SPARQL query: "Is there anyone who could be hurt?"

```
SELECT ?b WHERE { ?b
 <http://www.w3.org/1999/02/22
 -rdf-syntax-ns#type>
 <http://test.org/I_0_0.owl#Hit>.}
```

Applying similar techniques, the subsequent rules are formalized based on the descriptions outlined in the Risk Scenarios section:

$$Rule_2: Person(?x) \wedge age(?x,?b) \wedge greaterThan(?b,64) \wedge kitchen(?z) \wedge exist\_in(?x,?z) \wedge Sharp\_tool(?c) \wedge Furniture(?d) \wedge On\_Contact(?x,?d) \wedge On\_Contact(?c,?d) \rightarrow Hurt(?x)$$

$$Rule_3: Person(?x) \wedge Child(?x) \wedge Sharp\_tool(?c) \wedge On\_Contact(?x,?c) \rightarrow Hurt(?x,?c)$$

Rule 2 delineates the probability of an individual x_c, designated as $Elder_{Cp}$ or $Child_{Cp}$, to get $Hurt_{Cp}$ and subsequently be reclassified into that class if the rule body is true and satisfied. The body signifies an indirect association between the danger tool and x. Conversely, Rule 3 illustrates a direct relationship.

$$Rule_4: Unsafe\_Outdoor(?a) \wedge Person(?x) \wedge age(?x,?b) \wedge greaterThan(?b,64) \wedge exist\_in(?x,?a) \rightarrow being\_in\_dangerous\_place(?x)$$

$$Rule_5: Unsafe\_Outdoor(?a) \wedge Person(?x) \wedge age(?x,?b) \wedge lessThan(?b,13) \wedge exist\_in(?x,?a) \rightarrow being\_in\_dangerous\_place(?x)$$

Rule 4 and Rule 5 specify whether the individual x_c identified as an $Elder_{Cp}$ or a $Child_{Cp}$ is susceptible to the risk of $being\_in_dangerous\_place_{Cp}$, indicating that x_c is situated in hazardous outdoor or indoor environments.

$$Rule_6: Person(?x) \wedge age(?x,?b) \wedge greaterThan(?b,64) \wedge Mean\_of\_transport(?c) \wedge near(?x,?c) \wedge street(?d) \wedge under(?d,?x) \rightarrow Hit(?x)$$

$$Rule_7: Person(?x) \wedge age(?x,?b) \wedge greaterThan(?b,64) \wedge Mean\_of\_transport(?c) \wedge ride(?x,?c) \rightarrow Hit(?x)$$

Rule 6 states that if x_c is identified as an $Elder_{Cp}$ or a $Child_{Cp}$ and is situated near or around a $Mean\_of\_transport_{Cp}$, specifically on the street but not inside the means of transportation, then he will be reclassified as a Hit_{Cp} if the specified body is true. On the other hand, Rule 7 addresses the scenario where an $Elder_{Cp}$ is riding a $Mean\_of\_transport_{Cp}$.

$$Rule_8: Person(?x) \wedge age(?x,?b) \wedge greaterThan(?b,64) \wedge kitchen(?z) \wedge exist\_in(?x,?z) \wedge Hot\_tool(?c) \wedge On\_Contact(?x,?c) \rightarrow Burn(?x)$$

Finally, Rule 8 outlines the Burn risk scenario, wherein if x_c is classified as an $Elder_{Cp}$ and is either in contact with or in proximity to a Hot_tool_{Cp}, then he will be reclassified as a $Burn\_Cp$.

4 Study cases, experiments and tests

The motivation for this real-time healthcare system is twofold: Firstly, increasing statistics [1, 2] highlight a rise in accidents involving both the elderly and the young. Secondly, the system addresses the challenge of caring for elders and children amid busy schedules, where continuous supervision may be lacking. To evaluate its effectiveness, four typical risk scenarios are chosen: "Hurt," "Burn," "Existing in Dangerous Places," and "Hit."

4.1 Visual information extraction

The visual information extraction module is processed using YOLOv5, IoU, GCD, and ResNet18 to generate low/medium-level semantic descriptions in a three-triple format. These descriptions are then used to instantiate individuals in the Risks-Identification-Onto, which in turn passes to the high-level semantic risk identification.

The datasets used include the Charades dataset [37], the Actor-Action Dataset [38], and collected surveillance videos from YouTube.

4.1.1 Charades dataset

The Charades dataset [37] comprises 9,848 videos of typical indoor activities with an average runtime of 30 seconds and interactions with 46 object classes across 15 different interior environments. The dataset also includes a vocabulary of 30 verbs, translated into 157 action classes. Each video has various free-text annotations, action labels, action intervals, and classifications of interacting objects. Additionally, the dataset contains 27,847 textual descriptions of the videos and 41,104 labels for the 46 object classes, and it is divided into 7,986 training videos and 1,863 validation videos.

4.1.2 Actor-action dataset

The Actor-Action Dataset (A2D) [38] is used to identify actors and actions in videos at the same time. 8 action classes (climb, crawl, eat, fly, leap, roll, run, and walk) and 7 actor classes (adult, baby, ball, bird, car, cat, and dog) are included in A2D. It contains 3,782 videos, with 99 occurrences of each valid ”actor-action” tuple.

4.2 Risks-identification-onto

The ontology is constructed using Python 3.8 along with the ”owlready2” module and Protégé 5.5.0 with the OntoGraf plugin. ”owlready2” is an ontology-oriented programming module with robust capabilities for expressing and manipulating formal ontologies, along with agility for executing object-oriented programs, which is not possible with the mere use of ontology editors. This module includes parsers for the Web Ontology Language (OWL) and a quadstore for the Resource Description Framework (RDF) format (subject, property, object) [33]. Protégé is a free and open-source ontology editor, while OntoGraf is a Protégé plugin that explicitly displays the ontology entities.

The Hermit and Pellet [39] reasoners are used to check the consistency of the constructed Risks-Identification-Onto and to infer new knowledge regarding concepts, data properties, object properties, and individuals (Figure10).

4.3 Risk identification

The SWRL is used to generate rules based on the stated risk scenarios, using the Python module ”owlready2”, where risks can be derived and detected by reasoning this ”Rule Base” over the ”Fact Base,” as shown in Figure.11. Results show that the $Elder(Cp)$ has an indirect relationship (R) with the ”$knife2$ (C)” and is placed $near(R)$ the $oven/stove(C)$, therefore this Person could get *hurt* and *burned*.

Figure.12 illustrates a visualization of the entire process over the example case of Figure.4. The figure shows that the proposed real-time system can assist busy parents and caregivers in safeguarding elders and children by deducing potential risks in various scenarios with minimal inputs.

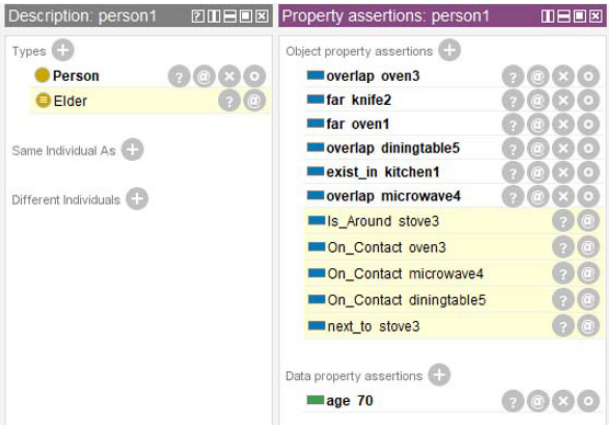

Figure 10: Pellet Inferences, the new information is highlighted with yellow.

Figure 11: Result of risk identification over the given scene Figure.4 and Figure.10

Additionally, as depicted in the ”Scene description generation” step, the system provides contextualized information with auto-generated semantic descriptions and alarms to highlight potential dangers.

The evaluation metrics used for the ”Real-time Healthcare system” with both Charades and A2D datasets are accuracy, precision, recall, and F1score. Table 4 presents the evaluation with the Charades dataset, while Table 5 is for A2D. Results show the efficiency of the proposed system in terms of risk identification with the Accuracy (Charades/A2D) $(98,29\%/99,43\%)$ for the hurt, $(97,61\%/97,61\%)$ for the Hit, $(97,34\%/98,40\%)$ for the burn, and $(97,41\%/97.25\%)$ for the Dangerous place. Precision (Charades/A2D) is $(97,78\%/98,89\%)$ for the hurt, $(97,12\%/96,15\%)$ for the Hit, $(97,89\%/98,95\%)$ for the burn, and $(96,36\%/98.15\%)$ for the Dangerous place. Finally, Recall ranges from $96,99\%$ to 98.88% for Charades and from $96,36\%$ to 100% for A2D, while F1score ranges from $97,25\%$ to 98.32% for Charades and from $97,39\%$ to $99,44\%$ for A2D.

The accuracy and calculated error rate of risk assignments according to each Cp and C are presented in the chart in Figure.13. The system can identify four types of risks, including ”Hit,” ”Hurt,” ”Burn,” and ”Dangerous Place,” with high accuracy and a low error rate. Furthermore, the confusion matrix (Figure. 14) depicts the predicted labels

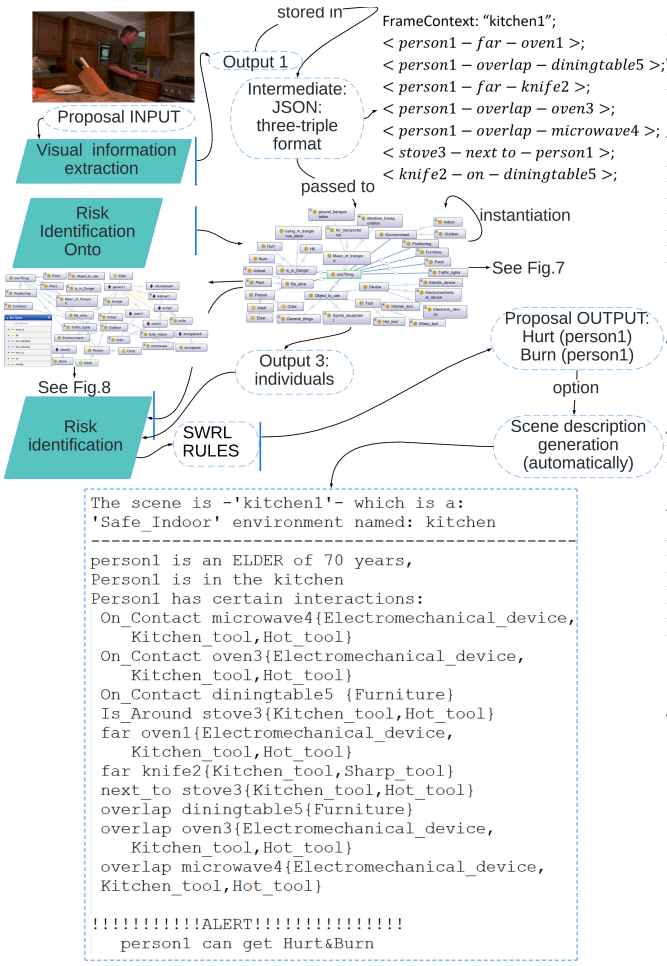

Figure 12: Risk identification process over Figure.4 .

based on the true risk labels.

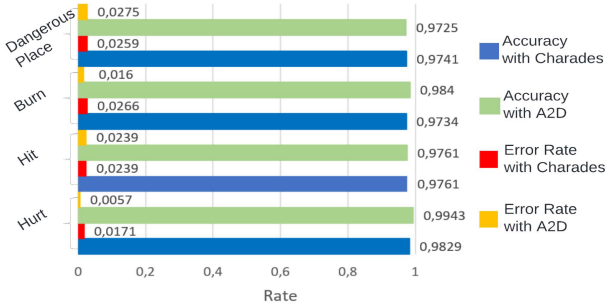

Figure 13: Performance of the proposal in terms of accuracy and error rate.

Table 6 compares the proposed approach to the works presented in [19], [4], and [10] across the following criteria: (1) ability to identify risks proactively, (2) the inference type, being deductive or inductive, (3) requirement for additional training or new datasets when adding new risks, and (4) the inference time. The proposal stands out for its

Table 4: Performance of the proposal using Charades [37].

Risk	Hurt	Hit	Burn	Dangerous _Place
Accuracy	0.9829	0.9761	0.9734	0.9741
Precision	0.9778	0.9712	0.9789	0.9636
Recall	0.9888	0.9806	0.9688	0.9815
F1score	0.9832	0.9758	0.9738	0.9725
Samples	175	209	188	116

Table 5: Performance of the proposal using A2D [38].

Risk	Hurt	Hit	Burn	Dangerous _Place
Accuracy	0.9943	0.9761	0.9840	0.9725
Precision	0.9889	0.9615	0.9895	0.9815
Recall	1.0000	0.9901	0.9792	0.9636
F1score	0.9944	0.9756	0.9843	0.9739
Samples	175	209	188	116

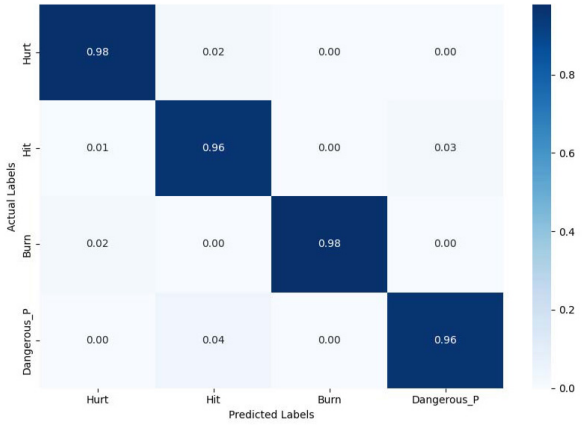

Figure 14: Confusion matrix of Hurt, Hit, Burn, Dangerous_Place

proactive risk identification using a combination of deductive and inductive reasoning, its ability to quickly adapt to new risks without additional training data, and real-time inference capabilities.

Table 7 compares the semantic-level capabilities of the proposed system to those of the previously mentioned approaches [19], [4], and [10]. The comparison criteria include: (1) the ability to achieve a high level of semantic understanding, (2) the provision of information-level semantic understanding, (3) the ability to deduce and reach a semantic knowledge level, (4) the capability to infer new knowledge distinct from the input data, and (5) the ability to provide an auto-generated semantic description. The proposed system demonstrates proficiency across all these criteria.

Table 6: Comparison between the proposal with the proposed approaches in [19], [4], and [10], where ”R” denotes ”Required.”

Criteria	Risk Identification	Deduction	Induction	training	New Data	Inference time
Proposal	√	√	√	×	×	real-time
[4]	×	×	√	R	R	2.5 ∼ 5 s
[10]	√	√	√	R	R	Not defined
[19]	√	×	√	R	R	real-time

Figure.15 showcases various inferences of risk and no-risk scenarios, including scenarios such as a) an elder on a cliff; b) an adult in the kitchen surrounded by several Electromechanical tools; c) an elder in the street and around means of transportation; and d) a child in the kitchen holding a knife. The results demonstrate the efficiency and consistency of the proposal in detecting elders and children, identifying and deducing risks before their occurrence, and generating a high-level semantic description that presents the risk type in real-time, with time intervals ranging from 0.13 to 0.29 seconds per frame. The risk identification system demonstrates real-time performance achieved through a combination of optimization techniques. For instance, it takes advantage of frameworks such as PyTorch, which is optimized for performance on both CPU and GPU architectures, as well as specific deep learning models such as YOLOv5 and GCD, along with logical reasoning integrated with ontology. These optimizations ensure efficient processing, making the system appropriate for real-time identification in a variety of scenarios.

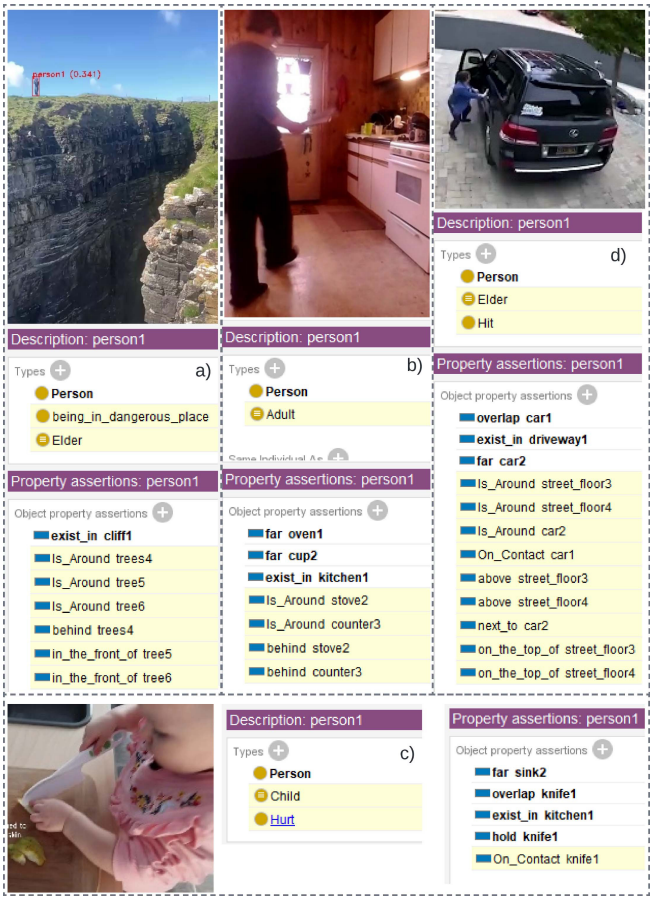

Figure 15: Four examples of the proposal’s results. a) Being in a dangerous place: An elder on a cliff, i.e., an unsafe environment. b) Safe: An adult can manage in the kitchen on his own. c) Hit An elder on the street/floor and around means of transportation is a car. d) Hurt: A child in the kitchen is in contact with a sharp tool, i.e., holding a knife.

5 Conclusion

This paper proposes a real-time healthcare semantic system that identifies visual dangers in surveillance videos for elders and children. The idea is to combine formal techniques and artificial vision. The approach consists of three modules: visual information extraction, ontology modeling, and risk detection. Each module is further subdivided into two bases: ”Fact Base” and ”Rule Base.” The Fact Base is generated using both extracted visual information and the newly constructed Risks-Identification-Ontology as well as its instantiations. Accordingly, the rule base is constructed using FOL and DL using the four frequent risk scenarios: ”Hurt,” ”Burn,” ”Existing in dangerous places,” and ”Hit.”

The risk identification process is achieved through reasoning using formal rules over the low/medium semantic outputs of data-driven approaches, which are mapped to a three-triple format. The Pellet and Hermit reasoners are used to perform the reasoning to identify and infer high-level semantic knowledge about risky situations, as well as to check the coherence of the ontology.

The proposed real-time system was tested on a variety of risky and safety cases. The results obtained demonstrate the efficiency of our proposed system, where it was successfully identified for each person on the scene in real-time with minimal use of resources and information. Moreover, it can automatically generate a semantic description. The efficiency of the system was tested using the very known and new datasets, i.e., the Charades dataset, the Actor-Action Dataset, and collected surveillance videos. The system gives an accuracy (Charades/A2D) $(98,29\%/99,43\%)$ for the Hurt, $(97,61\%/97,61\%)$ for the Hit, $(97,34\%/98,40\%)$ for the

Table 7: Comparison between the semantic-level of the proposal with the semantic-level of the approaches [19], [4], and [10].

Criteria	High-level semantic	Information	Knowledge	deduction	Semantic description
The proposal	√	√	√	√	√
[4]	×	√	×	×	×
[10]	√	√	√	×	×
[19]	×	√	×	×	×

Burn, and (97,41%/97.25%) for the Dangerous place, with a low error rate of 0,57% to 2,75%. Finally, compared with other approaches, the proposal can infer risks proactively in real-time, as well as deduce high-level semantic knowledge that differs from the inputted data.

Future expansions of this work include considering probabilistic scenarios to treat uncertainty in the generation of formal rules, incorporating advanced Machine Learning and Deep Learning (e.g., OpenPose), refining and expanding the ontology to cover a wider spectrum of concepts and domains, and extending the applicability of the system to diverse populations and application domains (e.g., risks identification in civil engineering and sites of construction).

Acknowledgment

The authors acknowledge the financial support and encouragement of the Research Laboratory on Computer Science's Complex Systems (ReLaCS2).

References

[1] Congxing Shi, Xiao Lin, Tingyuan Huang, Kai Zhang, Yanan Liu, Tian Tian, Pengyu Wang, Shimin Chen, Tong Guo, Zhiqiang Li, et al. "The association between wind speed and the risk of injuries among preschool children: New insight from a sentinel-surveillance-based study". In: *Science of the total environment* 856 (2023), p. 159005. https://doi.org/10.1016/j.scitotenv.2022.159005.

[2] Rik Dawson, Annie Feng, Juliana S Oliveira, Leanne Hassett, Catherine Sherrington, and Marina B Pinheiro. "Monitoring falls in residential aged care facilities: Agreement between falls incident reports and progress notes". In: *Australasian journal on ageing* (2024). https://doi.org/10.1111/ajag.13276.

[3] Takumi Ohnuki, Toru Abe, and Takuo Suganuma. "A Visual Monitoring Method for Infants in a Room". In: *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*. IEEE. 2020, pp. 258–259. https://doi.org/10.1109/gcce50665.2020.9292074.

[4] Peng-Jie Wang, Shao-Fu Lien, and Ming-Sui Lee. "A learning-based prediction model for baby accidents". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2019, pp. 629–633. https://doi.org/10.1109/icip.2019.8803820.

[5] Shajulin Benedict. "IoT-Enabled Remote Monitoring Techniques for Healthcare Applications–An Overview". In: *Informatica* 46.2 (2022). https://doi.org/10.31449/inf.v46i2.3912.

[6] Omar Doukari, James Wakefield, Pablo Martinez, and Mohamad Kassem. "An ontology-based tool for safety management in building renovation projects". In: *Journal of Building Engineering* 84 (2024), p. 108609. https://doi.org/10.1016/j.jobe.2024.108609.

[7] Satoshi Nishimura, Shusaku Egami, Takanori Ugai, Mikiko Oono, Koji Kitamura, and Ken Fukuda. "Ontologies of action and object in home environment towards injury prevention". In: *The 10th International Joint Conference on Knowledge Graphs*. 2021, pp. 126–130. https://doi.org/10.1145/3502223.3502239.

[8] Adel Zga and Brahim Nini. "Visual relationship extraction in images and a semantic interpretation with ontologies". In: *International Journal of Intelligent Information and Database Systems* 15.2 (2022), pp. 223–247. https://doi.org/10.1504/ijiids.2021.10041280.

[9] Shi Chen, Kazuyuki Demachi, and Feiyan Dong. "Graph-based linguistic and visual information integration for on-site occupational hazards identification". In: *Automation in Construction* 137 (2022), p. 104191. https://doi.org/10.1016/j.autcon.2022.104191.

[10] Yange Li, Han Wei, Zheng Han, Nan Jiang, Weidong Wang, and Jianling Huang. "Computer Vision-Based Hazard Identification of Construction Site Using Visual Relationship Detection and Ontology". In: *Buildings* 12.6 (2022), p. 857. https://doi.org/10.3390/buildings12060857.

[11] Malak Belkebir, Toufik Messaoud Maarouk, and Brahim Nini. "Integrating Ontology with Imaging and Artificial Vision for a High-Level Semantic: A Review". In: *Proceedings of the 2nd International*

Conference on Emerging Technologies and Intelligent Systems: ICETIS 2022, Volume 2. Springer. 2022, pp. 32–41. https://doi.org/10.1007/978-3-031-20429-6_4.

[12] Huma Parveen, Syed Wajahat Abbas Rizvi, and Raja Sarath Kumar Boddu. "Fuzzy-Ontology Based Knowledge Driven Disease Risk Level Prediction with Optimization Assisted Ensemble Classifier". In: *Data & Knowledge Engineering* (2024), p. 102278. https://doi.org/10.1016/j.datak.2024.102278.

[13] Cheng Zeng, Timo Hartmann, and Leyuan Ma. "ConSE: An ontology for visual representation and semantic enrichment of digital images in construction sites". In: *Advanced Engineering Informatics* 60 (2024), p. 102446. https://doi.org/10.1016/j.aei.2024.102446.

[14] Mourad Ellouze and Lamia Hadrich Belguith. "Semantic analysis based on ontology and deep learning for a chatbot to assist persons with personality disorders on Twitter". In: *Behaviour & Information Technology* (2023), pp. 1–20. https://doi.org/10.1080/0144929x.2023.2272757.

[15] Pim Borst and Hans Akkermans. "An ontology approach to product disassembly". In: *Knowledge Acquisition, Modeling and Management: 10th European Workshop, EKAW'97 Sant Feliu de Guixols, Catalonia, Spain October 15–18, 1997 Proceedings 10*. Springer. 1997, pp. 33–48. https://doi.org/10.1007/bfb0026776.

[16] Ahmad Abusukhon. "IOT Bracelets for Guiding Blind People in an Indoor Environment". In: *Journal of Communications Software and Systems* 19.2 (2023), pp. 114–125. https://doi.org/10.24138/jcomss-2022-0160.

[17] Sid Ahmed Hadri and Abdelkrim Bouramoul. "Friendy: A Deep Learning based Framework for Assisting in Young Autistic Children Psychotherapy Interventions". In: *Journal of Communications Software and Systems* 19.1 (2023), pp. 30–38. https://doi.org/10.24138/jcomss-2022-0074.

[18] Ming Li, Hui Dong, Fei Zhang, and Xiaoxiao Liu. "A Method for Top View Pedestrian Flow Detection Based on Small Target Tracking". In: *Informatica* 48.11 (2024). https://doi.org/10.31449/inf.v48i11.6033.

[19] Nelson RP Rodrigues, Nuno MC da Costa, César Melo, Ali Abbasi, Jaime C Fonseca, Paulo Cardoso, and João Borges. "Fusion Object Detection and Action Recognition to Predict Violent Action". In: *Sensors* 23.12 (2023), p. 5610. https://doi.org/10.20944/preprints202304.1242.v1.

[20] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788. https://doi.org/10.1109/cvpr.2016.91.

[21] Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma. "A Review of Yolo algorithm developments". In: *Procedia Computer Science* 199 (2022), pp. 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135.

[22] Xiaohang Shi, Jun Hu, Xueyue Lei, and Shiyou Xu. "Detection of flying birds in airport monitoring based on improved YOLOv5". In: *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*. IEEE. 2021, pp. 1446–1451. https://doi.org/10.1109/icsp51882.2021.9408797.

[23] Zexuan Guo, Chensheng Wang, Guang Yang, Zeyuan Huang, and Guo Li. "Msft-yolo: Improved yolov5 based on transformer for detecting defects of steel surface". In: *Sensors* 22.9 (2022), p. 3467. https://doi.org/10.3390/s22093467.

[24] Glenn Jocher, Alex Stoken, Jirka Borovec, Liu Changyu, Adam Hogan, Laurentiu Diaconu, Francisco Ingham, Jake Poznanski, Jiacong Fang, Lijun Yu, et al. "ultralytics/yolov5: v3. 1-bug fixes and performance improvements". In: *Zenodo* (2020). https://doi.org/10.5281/zenodo.4154370.

[25] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. "Places: A 10 million image database for scene recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 40.6 (2017), pp. 1452–1464. https://doi.org/10.1109/tpami.2017.2723009.

[26] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. "Generalized intersection over union: A metric and a loss for bounding box regression". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 658–666. https://doi.org/10.1109/cvpr.2019.00075.

[27] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. "Visual genome: Connecting language and vision using crowdsourced dense image annotations". In: *International journal of computer vision* 123 (2017), pp. 32–73. https://doi.org/10.1007/s11263-016-0981-7.

[28] Markos Diomataris, Nikolaos Gkanatsios, Vassilis Pitsikalis, and Petros Maragos. "Grounding consistency: Distilling spatial common sense for precise visual relationship detection". In: *Proceedings of the IEEE/CVF International Conference on Computer*

Vision. 2021, pp. 15911–15920. https://doi.org/10.1109/iccv48922.2021.01561.

[29] Gruber Tom. "Toward principles for the design of ontologies used for knowledge sharing". In: *Int. Workshop on Formal Ontology, 1993*. 1993. https://doi.org/10.1006/ijhc.1995.1081.

[30] Mike Uschold and Michael Gruninger. "Ontologies: Principles, methods and applications". In: *The knowledge engineering review* 11.2 (1996), pp. 93–136. https://doi.org/10.1017/s0269888900007797.

[31] Hannah A Valantine and Francis S Collins. "National Institutes of Health addresses the science of diversity". In: *Proceedings of the National Academy of Sciences* 112.40 (2015), pp. 12240–12242. https://doi.org/10.1073/pnas.1515612112.

[32] Michael R Genesereth and Nils J Nilsson. "Logical foundations of Artificial Intelligence". In: *New York: Morgan Kaufmann Publishers* (1987). https://doi.org/10.1016/c2009-0-27551-9.

[33] Lamy Jean-Baptiste and Lamy Jean-Baptiste. "The Python language: Adopt a snake!" In: *Ontologies with Python: Programming OWL 2.0 Ontologies with Python and Owlready2* (2021), pp. 9–48. https://doi.org/10.1007/978-1-4842-6552-9_2.

[34] Markus Krötzsch, Frantisek Simancik, and Ian Horrocks. "A description logic primer". In: *arXiv preprint arXiv:1201.4089* (2012). https://doi.org/10.48550/arXiv.1201.4089.

[35] Martin J O'Connor, Ravi D Shankar, Mark A Musen, Amar K Das, and Csongor Nyulas. "The SWRLAPI: A Development Environment for Working with SWRL Rules." In: *OWLED*. 2008.

[36] Martin J O'Connor and Amar K Das. "SQWRL: a query language for OWL." In: *OWLED*. Vol. 529. 2009, pp. 1–8.

[37] Gunnar A Sigurdsson, Gül Varol, Xiaolong Wang, Ali Farhadi, Ivan Laptev, and Abhinav Gupta. "Hollywood in homes: Crowdsourcing data collection for activity understanding". In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer. 2016, pp. 510–526. https://doi.org/10.1007/978-3-319-46448-0_31.

[38] Chenliang Xu, Shao-Hang Hsieh, Caiming Xiong, and Jason J Corso. "Can humans fly? action understanding with multiple classes of actors". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 2264–2273. https://doi.org/10.1109/cvpr.2015.7298839.

[39] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. "Pellet: A practical owl-dl reasoner". In: *Journal of Web Semantics* 5.2 (2007), pp. 51–53. https://doi.org/10.2139/ssrn.3199351.