Semantically Guided Scene Generation via Contextual Reasoning and Algebraic Measures

Loris Bozzato[1], Thomas Eiter[2], Rafael Kiesel[2] and Daria Stepanova[3]

[1]Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy
[2]Institute of Logic and Computation, Technische Universität Wien, Favoritenstraße 9-11, A-1040 Vienna, Austria
[3]Bosch Center for Artificial Intelligence, Renningen, Germany

Abstract

We recently presented the MR-CKR framework to reason with knowledge overriding across contexts organized in multi-relational hierarchies. Reasoning is realized via ASP with Algebraic Measures, allowing for flexible definitions of preferences. In this paper, we show how to apply our theoretical work to autonomous-vehicle scene data: we apply MR-CKR to the problem of generating challenging scenes for autonomous vehicle learning. In practice, most of the scene data for AV learning models common situations, thus it might be difficult to capture cases where a particular situation occurs (e.g. partial occlusions of a crossing pedestrian). The MR-CKR model allows for data organization exploiting the multi-dimensionality of such data (e.g., temporal and spatial dimension). Reasoning over multiple contexts enables the verification and configuration of scenes, using the combination of different scene ontologies. We describe a framework for semantically guided data generation, based on a combination of MR-CKR and algebraic measures. The framework is implemented in a proof-of-concept prototype exemplifying some cases of scene generation.

1. Introduction

Testing and evaluation are important steps in the development and deployment of Automated Vehicles (AVs). To comprehensively evaluate the performance of AVs, it is crucial to test the AVs' perception systems in safety-critical scenarios, which rarely happen in naturalistic driving environment, but still possible in practice. Therefore, the targeted and systematic generation of such corner cases becomes an important problem.

Most existing studies focus on generating adversarial examples for perception systems of AVs which are concerned with very simple perturbations in the input (e.g., changing the color or position of a vehicle). More in general, the generation of adversarial or challenging examples for neural models is an important problem that gained interest both in industry[1]

ASPOCP 2023, 16th Workshop on Answer Set Programming and Other Computing Paradigms, July 9 - 15, 2023, London
bozzato@fbk.eu (L. Bozzato); thomas.eiter@tuwien.ac.at (T. Eiter); rafael.kiesel@web.de (R. Kiesel); Daria.Stepanova@de.bosch.com (D. Stepanova)
https://dkm.fbk.eu/author/lorisbozzato/ (L. Bozzato); http://www.kr.tuwien.ac.at/staff/eiter/ (T. Eiter); https://raki123.github.io/ (R. Kiesel); https://www.bosch-ai.com/about-us/our-people/daria-stepanova/ (D. Stepanova)
0000-0003-1757-9859 (L. Bozzato); 0000-0001-6003-6345 (T. Eiter); 0000-0002-8866-3452 (R. Kiesel); 0000-0001-8654-5121 (D. Stepanova)

[1]https://www.efemarai.com/

and research [1, 2, 3, 4]. Generation of inputs is mostly performed by numerical methods. For example, [1] uses small numerical perturbations of images, [2] uses an optimization that minimizes the numerical change of the input data such that it leads to a different prediction of the network, and [4] generates adversarial text for natural language processing by performing minimal replacements of characters.

On the other hand, limited efforts have been put on the symbolic generation of complex scenes: this is the problem we want to consider in this paper, in particular aiming at an ontology-based and context-specific scene generation (e.g., child walking a dog in the evening in rainy weather). Specifically, we define our task of interest as follows: given an existing scene (represented by a scene graph) from a known dataset, we want to generate a new set of scenes that are variations of the current scene and are:

1. **Realistic:** consistent with the ontologies describing objects in the scene (e.g., traffic signs usually do not move);
2. **Interesting:** they satisfy a semantic restriction, which tells us that the scene is for example, "dangerous" or challenging for our prediction model (e.g., seeing a cat in the middle of the street requires special action);
3. **Similar:** changing the original scene to the generated scenes requires only small variations.

We propose to use symbolic methods to generate valid and challenging scenes on the base of existing scene graphs and semantic definitions of scenes. In particular, *Multi-Relational Contextualized Knowledge Bases (MR-CKRs)* [5, 6] are a useful formalism for this: here, ontological knowledge is contextualized such that in different contexts it may have different interpretations (possibly with non-monotonic effects). Thus, MR-CKRs can help us in generating *realistic* scenes, since they are capable of handling the background ontologies describing the AV domain. Additionally, we may have different contextualized notions of *interestingness*. E.g., it may be that a dog is by default not considered dangerous but in a special context, next to a crowded street, it is. MR-CKRs also allow us to express this by associating different independent semantic restrictions on scenes within different contexts. Another benefit of MR-CKRs is that they come with a translation to Answer Set Programming (ASP), which can be used to easily express and efficiently solve hard logical problems.

Additionally, we need a way to measure how similar the generated scenes are to the original scene that we started from: *Algebraic Measures* [7] are of great use here. They are a general framework from the field of ASP that allows us to measure quantities associated with solutions. As such, they are also capable of expressing a similarity measure of scenes, which we need.

Our main contributions can be summarized as follows:

- We provide a novel framework for semantically guided data generation, which can be adversarial or training data. An overview of the framework is presented in Section 2. By basing our framework on a combination of MR-CKRs and Algebraic Measures we obtain a highly flexible approach with efficient solving options by employing translations to ASP.
- MR-CKRs allow us (i) to incorporate ontological background knowledge ensuring realism of the generated data and (ii) to contextualize the notion of what makes a generated input interesting. The MR-CKR formalism and its use are presented in Section 3.

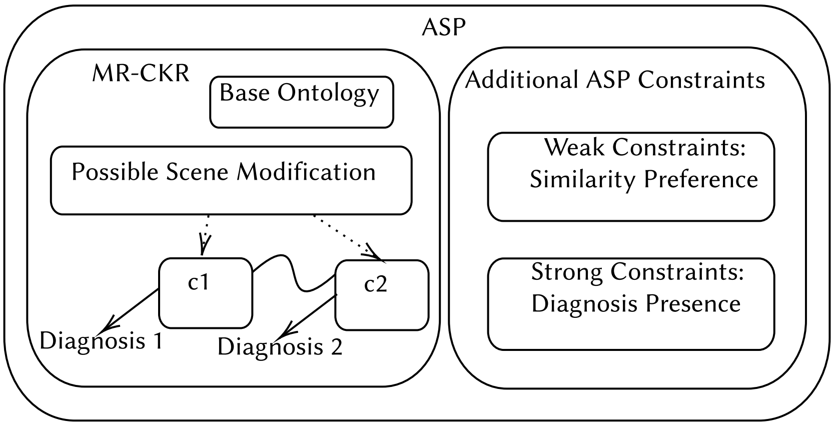

Figure 1: The general framework for generating (similar) interesting scenes with ASP according to (possibly) related diagnoses defined by an MR-CKR.

- Algebraic Measures enable the maximization of similarity between original and generated data. The definition of similarity by Algebraic Measures and its realization via weak constraints is presented in Section 4.
- We provide a proof of concept implementation of our framework in ASP. Our prototype (presented in Section 5) for scene generation in the domain of AV is intentionally kept minimal but shows promise.

2. Framework Overview

Before we go into the technical details of how we generate descriptions of new challenging training scenes, we provide a general structural overview of our framework.

The architecture of our framework is described in Figure 1. The schema shows that we use an MR-CKR to define, on the one hand, the possible scene modifications and on the other hand different contexts, here $c1$ and $c2$, that specify possibilities for a scene to be challenging respectively interesting. Ideally, these different diagnoses correspond to insights of a neural network engineer for poor performance of the current neural network. For example, in the AV context, we might observe that a car does not stop in the correct location when there is not only a stop sign but also a stop line marking that specifies where the car should stop. Here, we would therefore want to modify scenes in such a manner that they have both a stop sign and a stop line marking.

In general, the goal is to generate more scenes that we suspect the network also performs badly on, such that we have adversarial examples that we can use to train the neural network in the hope of improving its performance on these situations that are presumably hard for it, due to a lack of training data. Given the diagnoses in different contexts that may be related via specialization or otherwise, we can then use the ASP translation of an MR-CKR to obtain an equivalent ASP encoding to compute models, i.e., representations of generated scenes that are realistic according to the base ontology included in the MR-CKR.

Moreover, we add further ASP constraints to ensure that the modifications of the scene make the resulting scene (a) challenging, i.e., it satisfies a diagnosis (using strong constraints) and (b) close to a given starting scene (using weak constraints that express the algebraic measure).

With such combination, we can thus obtain realistic, challenging scenes that are as similar to the starting scene as possible. On top of that, the different contexts allow us to specify different types of target diagnoses resulting in one generated scene that includes it per context.

In the following, we substantiate our abstract idea by formalizing how we generate scenes with MR-CKR and measure their similarity with Algebraic Measures.

3. Formalization of the Scene Generation Problem in MR-CKR

We begin by briefly introducing the MR-CKR framework and then provide a solution for the scene generation task making use of MR-CKR.

3.1. MR-CKR: Multi-relational Contextualized Knowledge Bases

We summarize in this subsection the main definitions of the MR-CKR formalism from [6].

We assume the customary definitions for description logics (see, e.g., [8]), where we consider a generic description logic language $\mathcal{L}_\Sigma$ based on a DL signature Σ, which is composed of a set of concept names NC, role names NR and individual names NI.

The contextual structure of a MR-CKR is defined by a nonempty set $\mathbf{N} \subseteq \mathrm{NI}$ of *context names* and a set $\mathcal{R}$ of one or more *contextual relations* $\prec_i$ over them, which are strict (partial) orders $\prec_i \subseteq \mathbf{N} \times \mathbf{N}$. A way to define contextual relations is to use *contextual dimensions* [9, 10], that is a set of contextual "coordinates" associated to each of the contexts: in the case of scene descriptions, for example, these can represent the time of the day, location type or situation occurring in a scene.

With each context of $\mathbf{N}$, we associate a contextual language $\mathcal{L}_{\Sigma,\mathbf{N}}$ as an extension of $\mathcal{L}_\Sigma$ representing its local language; in particular, axioms inside contexts can be specified as defeasible (i.e., they can be overridden in case of exceptions) with respect to one of the contextual relations composing the contextual structure. Formally, an *r-defeasible axiom* is any expression of the form $\mathrm{D}_r(\alpha)$, where α is an axiom of $\mathcal{L}_\Sigma$ and $\prec_r \in \mathcal{R}$. Multi-relational CKRs are then composed by a global structure of context based on the contextual relations in $\mathcal{R}$ and a set of DL knowledge bases associated to each of the local contexts.

Definition 1 (multi-relational simple CKR). *A* multi-relational simple CKR (sCKR) *over* Σ *and* $\mathbf{N}$ *is a structure* $\mathfrak{K} = \langle \mathfrak{C}, \mathrm{K}_\mathbf{N} \rangle$ *where: (i).* $\mathfrak{C}$ *is a structure* $(\mathbf{N}, \prec_1, \ldots, \prec_m)$ *where each* $\prec_i$ *is a contextual relation over* $\mathbf{N}$*; (ii).* $\mathrm{K}_\mathbf{N} = \{\mathrm{K}_\mathsf{c}\}_{\mathsf{c} \in \mathbf{N}}$ *for each context name* $\mathsf{c} \in \mathbf{N}$, K_c *is a DL knowledge base over* $\mathcal{L}_{\Sigma,\mathbf{N}}$.

Example 1. *We provide a simple example of MR-CKR to better explain the intended use of defeasible axioms. Consider the sCKR* $\mathfrak{K} = \langle \mathfrak{C}, \{\mathrm{K}_1, \mathrm{K}_2\} \rangle$ *composed of the following elements:*

$$\begin{aligned} \mathfrak{C} = &\quad \{\mathsf{c}_2 \prec_c \mathsf{c}_1\}, \\ \mathrm{K}_1 = &\quad \{\mathrm{D}_c(\mathit{Dog} \sqsubseteq \neg \mathit{DangerousAnimal})\}, \\ \mathrm{K}_2 = &\quad \{\mathit{Dog} \sqsubseteq \mathit{DangerousAnimal}, \mathit{Dog}(d)\}. \end{aligned}$$

Intuitively, we want to ensure that in the more specific context c_2*, dogs are viewed as dangerous animals, thus the more general defeasible axiom in* c_1 *is not applied to the instance* d *of* Dog*.*

Interpretations of MR-CKRs $\mathfrak{I} = \{\mathcal{I}(\mathsf{c})\}_{\mathsf{c} \in \mathbf{N}}$ are families of DL interpretations associated to each of the contexts that agree on the interpretation of individuals. Given an axiom $\alpha \in \mathcal{L}_\Sigma$ with FO-translation $\forall \mathbf{x}.\phi_\alpha(\mathbf{x})$, we say that the *instantiation* of α with a tuple $\mathbf{e}$ of individuals in NI, written $\alpha(\mathbf{e})$, is the specialization of α to $\mathbf{e}$, i.e., $\phi_\alpha(\mathbf{e})$, depending on the type of α. Given a relation $\succeq_r$, we denote with $\succeq_{-r}$ any other relation in $\mathcal{R}$ different from $\succeq_r$.

A *clashing assumption* for a context c and relation r is a pair $\langle \alpha, \mathbf{e} \rangle$ such that $\alpha(\mathbf{e})$ is an axiom instantiation of α, and $\mathsf{c}' \succeq_{-r} \mathsf{c}'' \succ_r \mathsf{c}$. A *clashing set* for $\langle \alpha, \mathbf{e} \rangle$ is a satisfiable set S of ABox assertions s.t. $S \cup \{\alpha(\mathbf{e})\}$ is unsatisfiable. Intuitively, clashing assumptions represent the assumption that $\mathbf{e}$ are exceptional individuals of α, clashing sets provide a justification of such exceptionality. We extend interpretations with clashing assumptions in what we call CAS-interpretations: a *CAS-interpretation* is a structure $\mathfrak{I}_{CAS} = \langle \mathfrak{I}, \overline{\chi} \rangle$ where $\mathfrak{I}$ is an interpretation and $\overline{\chi} = \{\chi_1, \ldots, \chi_m\}$ such that each χ_i, for $i \in \{1, \ldots, m\}$, maps every $\mathsf{c} \in \mathbf{N}$ to a set $\chi_i(\mathsf{c})$ of clashing assumptions for context c and relation $\prec_i$.

We say that a CAS-interpretation $\mathfrak{I}_{CAS} = \langle \mathfrak{I}, \overline{\chi} \rangle$ is a *CAS-model* for $\mathfrak{K}$ (denoted $\mathfrak{I}_{CAS} \models \mathfrak{K}$), if it verifies all strict and defeasible axioms of $\mathfrak{K}$, i.e., informally: (i) $\mathcal{I}(\mathsf{c}') \models \alpha$ for every strict axiom α in a context c with $\mathsf{c}' \preceq_* \mathsf{c}$; (ii) $\mathcal{I}(\mathsf{c}') \models \alpha$ for every i-defeasible axiom $\mathrm{D}_i(\alpha)$ in a context c related by a non-i relation, that is $\mathsf{c}' \preceq_{-i} \mathsf{c}$; (iii) $\mathcal{I}(\mathsf{c}'') \models \phi_\alpha(\mathbf{d})$ for every i-defeasible axiom $\mathrm{D}_i(\alpha)$ in a context c with $\mathsf{c}'' \prec_i \mathsf{c}' \preceq_{-i} \mathsf{c}$ and $\mathbf{d}$ not exceptional, that is $\langle \alpha, \mathbf{d} \rangle \notin \chi_i(\mathsf{c}'')$.

We say that $\langle \alpha, \mathbf{e} \rangle \in \chi_i(\mathsf{c})$ is *justified* for a CAS model $\mathfrak{I}_{CAS}$, if some clashing set $S_{\langle \alpha, \mathbf{e} \rangle, \mathsf{c}}$ exists such that $\mathcal{I}'(\mathsf{c}) \models S_{\langle \alpha, \mathbf{e} \rangle, \mathsf{c}}$ (and for all $\mathfrak{I}'_{CAS}$ agreeing on the interpretation of individuals). A CAS model $\mathfrak{I}_{CAS}$ of a sCKR $\mathfrak{K}$ is *justified*, if every $\langle \alpha, \mathbf{e} \rangle \in \overline{\chi}$ is justified in $\mathfrak{K}$.

We provide a *local preference* on clashing assumption sets for each of the relations:

(LP). $\chi_i^1(\mathsf{c}) > \chi_i^2(\mathsf{c})$, if for every $\langle \alpha_1, \mathbf{e} \rangle \in \chi_i^1(\mathsf{c}) \setminus \chi_i^2(\mathsf{c})$ with $\mathrm{D}_i(\alpha_1)$ at a context $\mathsf{c}_1 \succeq_{-i} \mathsf{c}_{1b} \succ_i \mathsf{c}$, some $\langle \alpha_2, \mathbf{f} \rangle \in \chi_i^2(\mathsf{c}) \setminus \chi_i^1(\mathsf{c})$ exists with $\mathrm{D}_i(\alpha_2)$ at context $\mathsf{c}_2 \succeq_{-i} \mathsf{c}_{2b} \succ_i \mathsf{c}$ s.t. $\mathsf{c}_{1b} \succ_i \mathsf{c}_{2b}$.

Intuitively, $\chi_i^1(\mathsf{c})$ is preferred to $\chi_i^2(\mathsf{c})$ if $\chi_i^1(\mathsf{c})$ exchanges the "more costly" exceptions of $\chi_i^2(\mathsf{c})$ at more specialized contexts with "cheaper" ones at more general contexts. A *model preference* is obtained by combining the preferences of the relations: it is a global lexicographical ordering on models where each $\prec_i$ defines the ordering at the i-th position.

(MP). $\mathfrak{I}^1_{CAS} = \langle \mathfrak{I}^1, \chi_1^1, \ldots, \chi_m^1 \rangle$ is preferred to $\mathfrak{I}^2_{CAS} = \langle \mathfrak{I}^2, \chi_1^2, \ldots, \chi_m^2 \rangle$ if:

(i) there exists $i \in \{1, \ldots, m\}$ and some $\mathsf{c} \in \mathbf{N}$ s.t. $\chi_i^1(\mathsf{c}) > \chi_i^2(\mathsf{c})$ and not $\chi_i^2(\mathsf{c}) > \chi_i^1(\mathsf{c})$, and for no context $\mathsf{c}' \neq \mathsf{c} \in \mathbf{N}$ it holds that $\chi_i^1(\mathsf{c}') < \chi_i^2(\mathsf{c}')$ and not $\chi_i^2(\mathsf{c}') < \chi_i^1(\mathsf{c}')$.

(ii) for every $j < i \in \{1, \ldots, m\}$, it holds $\chi_j^1 \approx \chi_j^2$ (i.e., (i) or its converse do not hold for $\prec_j$).

Finally, we say that an interpretation $\mathfrak{I}$ is a *CKR model* of $\mathfrak{K}$ (in symbols, $\mathfrak{I} \models \mathfrak{K}$) if: (i) $\mathfrak{K}$ has some justified CAS model $\mathfrak{I}_{CAS}$; (ii) there exists no justified $\mathfrak{I}'_{CAS}$ that is preferred to $\mathfrak{I}_{CAS}$.

Example 2. *Using the semantics mechanism from above, we can show how to interpret the sCKR in the previous example. We can consider the CAS-interpretation* $\mathfrak{I}_{CAS} = \langle \mathfrak{I}, \overline{\chi} \rangle$ *where*

$\chi(\mathsf{c}_2) = \{\langle Dog \sqsubseteq \neg DangerousAnimal, d\rangle\}$. *This implies that $\mathfrak{I}_{CAS}$ is a CAS-model if the defeasible axiom of c_1 is not applied to the only Dog in c_2, as expected. Note that such CAS-model is also justified, since the clashing assumption admits the clashing set $\{Dog(d), DangerousAnimal(d)\}$: thus, considering that no other alternative CAS-model that is minimal with respect to the preference can be defined, the considered interpretation is also a CKR-model. The preference defined above is useful to prefer defeasible axioms in the most specific contexts: for example, if $Dog \sqsubseteq DangerousAnimal$ in c_2 was defined as defeasible (w.r.t. the same contextual relation of the above defeasible axiom), the context below c_2 would have preferred the more specific axiom.*

As a method to implement reasoning on MR-CKRs, we provided in [6] a translation for MR-CKRs to ASP logic programs, which can be used to reason on instance checking and query answering in a given context. It is based on a uniform encoding of DL knowledge bases using a materialization calculus, which is extended for defeasible axioms. To give a flavor of the translation, the following rules define the conditions for application and overriding of a defeasible inclusion like $D_{rel1}(E \sqsubseteq F)$:

```
instd(X,F,C,T) :- def_subClass(E,F,C1,REL1), instd(X,E,C,T),
                  prec(C,C2, REL1), preceq(C2,C1, REL2), REL1 != REL2,
                  not ovr(subClass,X,E,F,C1,C,REL1).

ovr(subClass,X,E,F,C1,C,REL1) :- def_subClass(E,F,C1,REL1),
                                 prec(C,C2,REL1), preceq(C2,C1,REL2),
                                 REL1 != REL2, instd(X,E,C,"main"),
                                 not test_fails(nlit(X,F,C)).
```

For details and discussion, we refer to [6].

3.2. Scene generation in MR-CKR

Following the intuitive structure of Figure 1, the role of MR-CKR in our architecture is to define the logical constraints of the scenes we want to generate, on the basis of a common scene ontology. Given its multi-contextual structure, the MR-CKR is useful to provide *(i)* a complex representation (a contextualization) of the contents of the base scene and *(ii)* its modifications towards the different diagnoses of interest.

With respect to the second aspect, the basic organization of contexts can be defined as in Figure 2. The contexts of this structure are related by a contextual relation $\succ_{sim}$, denoting the relation of similarity: the upper context $Exchange$ contains, in form of sim-defeasible axioms, the axioms that can be modified in the diagnosis scenes. In the $Base$ context, we assume to have the description of the base scene and the base axioms of the scene description ontology. The contexts $Diagnosis\text{-}1, \ldots, Diagnosis\text{-}N$ then provide the different modifications to the base scene that we are interested in modeling. The kind of axioms that are needed to model the different modifications depend on the kind of changes (additions, deletions, etc.) that we want to admit in scene modifications; more details on such axioms will be provided in the following sections, where we consider specific modifications.

With respect to the scene contextualization, we can take advantage of the multi-relational nature of MR-CKR to further define the properties of the context in which the scene takes place.

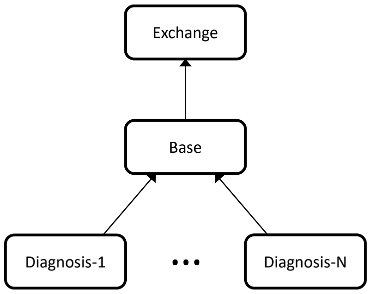

Figure 2: General structure of contexts for scene modification.

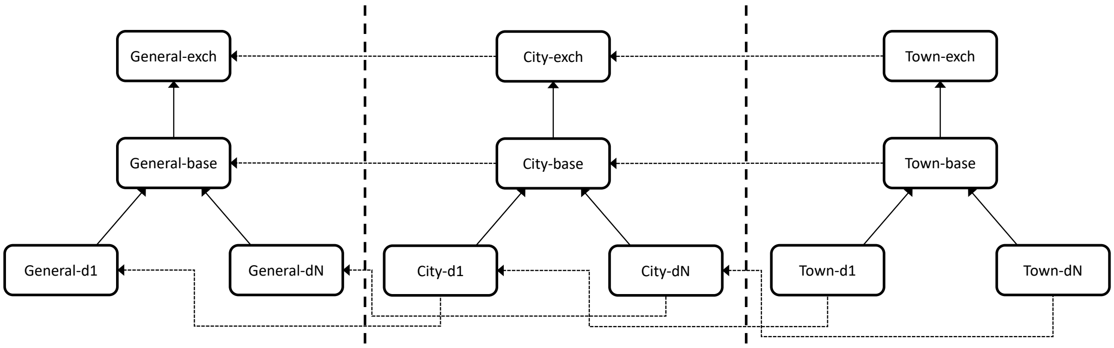

Figure 3: Example of multi-relational contextual structure for scene representation.

Example 3. *An example of such contextualization is shown in Figure 3. In this contextual structure, the relation given by the horizontal arrows represents the specialization of scenes with respect to the specificity of the location: starting from axioms that are verified for* general *scenes, we can add further logical constraints that are true for* city *scenes and then* town *scenes. Note that such direction is orthogonal to the base contextual structure described above: location-defeasible axioms can express knowledge that is, e.g., accepted in general locations, but not valid in town scenes.*

After modelling scenes by such framework, we want to use the translation of MR-CKR to ASP in order to generate the possible models of the diagnoses contexts: these models then correspond to alternative generated scenes. However, we now need a method to provide a measure for the *similarity* of the generated scenes with respect to the scenes of interest: as we detail in the following sections, this can be easily defined by means of algebraic measures.

4. Formalization of Similarity using Algebraic Measures

If we generate a new scene based on a starting scene, we want to optimize a measure of similarity. Here, we use algebraic measures [7], which measure and aggregate quantities associated with models by evaluating a formula over a *semiring*.

Definition 2 (Semiring). *A commutative* semiring $\mathcal{S} = (S, \oplus, \otimes, e_{\oplus}, e_{\otimes})$ *is an algebraic structure with binary infix operations* $\oplus, \otimes$ *such that*

1. $\oplus$ *and* $\otimes$ *are associative and commutative,*
2. $\otimes$ *right and left distributes over* $\oplus$*,*
3. $e_\oplus$ *(resp.* $e_\otimes$*) is a neutral element for* $\oplus$ *(resp.* $\otimes$*), and*
4. $e_\oplus$ *annihilates* S*, i.e.,* $\forall s \in S : s \otimes e_\oplus = e_\oplus = e_\oplus \otimes s$.

Examples of well-known commutative semirings are

- $\mathbb{F} = (\mathbb{F}, +, \cdot, 0, 1)$, where $\mathbb{F} \in \{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$, the semiring over the numbers in $\mathbb{F}$ with addition and multiplication,
- $\mathcal{P} = ([0, 1], +, \cdot, 0, 1)$, the probability semiring,
- $\mathcal{R}_{\min,+} = (\mathbb{R} \cup \{\infty\}, \min, +, \infty, 0)$, the min-plus semiring.

Another list of semirings, which is annotated with applications, can be found in [11].

In order to connect the quantitative aspects of semirings and the qualitative ones of logics we use weighted logics [12].

Definition 3 (Weighted Logic). *Let* $\mathcal{V}$ *be a set of propositional variables and let* $\mathcal{R} = (R, \oplus, \otimes, e_\oplus, e_\otimes)$ *be a semiring. A* weighted *(propositional) formula over* $\mathcal{R}$ *is of the form* α *given by the grammar*

$$\alpha ::= k \mid v \mid \neg v \mid \alpha + \alpha \mid \alpha * \alpha$$

where $k \in R$ *and* $v \in \mathcal{V}$*. The value* $[\![\alpha]\!]_\mathcal{R}(\mathcal{I})$ *of* α *w.r.t. an* interpretation $\mathcal{I} \subseteq \mathcal{V}$ *is defined as:*

$$\begin{aligned}
[\![k]\!]_\mathcal{R}(\mathcal{I}) &= k \\
[\![v]\!]_\mathcal{R}(\mathcal{I}) &= \begin{cases} e_\otimes & v \in \mathcal{I} \\ e_\oplus & \textit{otherwise.} \end{cases} \quad (v \in \mathcal{V}) \\
[\![\neg v]\!]_\mathcal{R}(\mathcal{I}) &= \begin{cases} e_\oplus & v \in \mathcal{I} \\ e_\otimes & \textit{otherwise.} \end{cases} \quad (v \in \mathcal{V}) \\
[\![\alpha_1 + \alpha_2]\!]_\mathcal{R}(\mathcal{I}) &= [\![\alpha_1]\!]_\mathcal{R}(\mathcal{I}) \oplus [\![\alpha_2]\!]_\mathcal{R}(\mathcal{I}) \\
[\![\alpha_1 * \alpha_2]\!]_\mathcal{R}(\mathcal{I}) &= [\![\alpha_1]\!]_\mathcal{R}(\mathcal{I}) \otimes [\![\alpha_2]\!]_\mathcal{R}(\mathcal{I}).
\end{aligned}$$

Algebraic measures combine the qualitative language of ASP with the quantitative one of weighted logic.

Definition 4 (Algebraic Measure). *An* algebraic measure $\mu = \langle \Pi, \alpha, \mathcal{R} \rangle$ *consists of an answer set program* Π*, a weighted formula* α*, and a semiring* $\mathcal{R}$*. Then, the weight of an answer set* $\mathcal{I} \in \mathcal{AS}(\Pi)$ *under* μ *is defined by*

$$\mu(\mathcal{I}) = [\![\alpha]\!]_\mathcal{R}(\mathcal{I}).$$

Intuitively, Π determines the solutions and α assigns them weights. Algebraic measures offer further queries that aggregate the weights of (some) solutions [7] that we however do not require here.

4.1. Similarity of Scenes

We measure the similarity of two scenes as the minimum cost required to modify one scene to be equal to another one. There are many ways to modify a scene. For simplicity, we restrict ourselves to deletion of object o from class c and addition of object o to class c, denoted $\mathrm{deletion}(c, o)$ and $\mathrm{addition}(c, o)$, with fixed costs[2] $\mathrm{cost}(Add)$ and $\mathrm{cost}(Del)$ respectively. Advanced options would be class variation, displacement, or property variation.

We use the weighted formula α_{cost}

$$\Pi_{\text{class } c, \text{individual } i}(\mathrm{addition}(c, i) * \mathrm{cost}(Add) + \neg\mathrm{addition}(c, i))$$
$$* \, \Pi_{\text{class } c, \text{individual } i}(\mathrm{deletion}(c, i) * \mathrm{cost}(Del) + \neg\mathrm{deletion}(c, i))$$

over the semiring $\mathcal{R}_{\min,+}$, which takes the minimum of different options and sums up the costs of atomic modifications in one option.

Example 4. *For example, consider the interpretation*

$$\mathcal{I} = \{\mathrm{addition}(RollingContainer, i_1), \mathrm{deletion}(Child, i_2), \mathrm{deletion}(Child, i_3)\}.$$

Intuitively, this means that we add the object i_1 to the concept $RollingContainer$ and remove the objects i_2 and i_3 from the concept $Child$. Thus, we expect a cost of $\mathrm{cost}(Add) + 2 \cdot \mathrm{cost}(Del)$.

Since the interpretation $\mathcal{I}$ contains $\mathrm{addition}(RollingContainer, i_1)$ *the first row of α_{cost} evaluates to* $\mathrm{cost}(Add)$ *and the second row evaluates to* $\mathrm{cost}(Del) \otimes \mathrm{cost}(Del)$. *Since we use $\mathcal{R}_{\min,+}$ the operation $\otimes$ is $+$ and we obtain* $\mathrm{cost}(Add) + 2 \cdot \mathrm{cost}(Del)$ *as the final cost, as expected.*

While algebraic measures are a useful and highly general tool, to specify quantitative measures for the answer sets of logic programs, most solvers for ASP currently do not support optimization of the weight of an algebraic measure. However, algebraic measures over the semiring $\mathcal{R}_{\min,+}$ can be translated to *weak constraints* [13].

Since we intentionally kept the measure simple, it is sufficient to use the weak constraints

$$:\sim \mathrm{addition}(C, I).[\mathrm{cost}(Add), add, C, I] \qquad :\sim \mathrm{deletion}(C, I).[\mathrm{cost}(Del), del, C, I]$$

which add a cost of $\mathrm{cost}(Add)$ for each individual I and concept C such that we add I to concept C (and anologously for deletion). In general, this is not as simple since measures can specify complex algebraic expression, which is why we prefer the specifiation via measures.

5. Implementation Prototype for Scene Generation in Autonomous Driving

We have implemented and tested our approach using an example from Autonomous Driving.[3] Here, we reconstructed and slightly extended the base ontology from [14, 15].[4] It features

[2] i.e., costs do not depend on the class or individual.
[3] https://github.com/raki123/MR-CKR
[4] https://github.com/boschresearch/ad_cskg

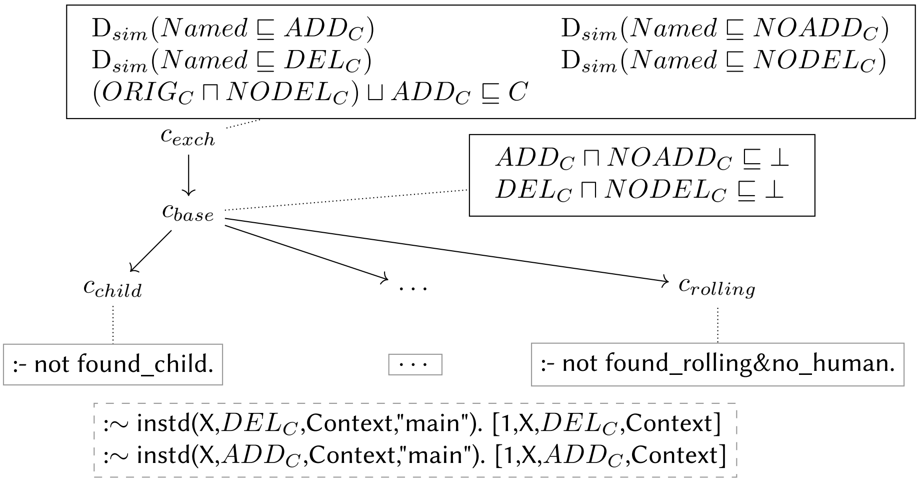

Figure 4: MR-CKR and additional constraints used in our prototype for autonomous driving.

different scenes, each annotated with the objects included in it. The included objects are annotated with information about them, such as their type. Furthermore, the ontology includes axioms that add knowledge about the relationship between different concepts in the ontology. Overall, the base ontology has more than 3 million axioms, concerning knowledge of 41 concepts, more than 100 scenes, and more than 50,000 objects.

We provide an overall sketch of the construction of the MR-CKR and its interplay with ASP constraints in Figure 4. We go over the different parts step by step.

MR-CKR Encoding. Here, we describe how scenes can be modified (annoted via solid black boxes in Figure 4). Namely, we state that every named object should either be added or not added (resp. deleted) to each modifiable concept. Additionally, we ensure that the modification takes effect by stating that a concept contains exactly those objects (i) which it originally contained before modification and that were not deleted, or (ii) that were added.

Apart from that, the MR-CKR contains the axioms from the base ontology and one context C_i for each diagnosis i. We use the following diagnoses:

1. c_{child}: the scene contains an object that is in the class $Child$. It is dangerous due to unpredictable behaviour compared to other humans.

2. $c_{rolling}$: the scene contains an object that is in the class $RollingContainer$ but no object in the class $Human$. It is dangerous due to unpredictable behaviour of the rolling container.

We then translate the resulting encoding to ASP using the CKRew software[5] from [6].

Additional ASP Constraints. To ensure that in context C_i the diagnosis i is derived, we add strong constraints (annoted via solid gray boxes in Figure 4). For example, for the first diagnosis in context C_1, we ensure the presence of a child using the rules

[5]https://github.com/dkmfbk/ckrew

```
found_child :- instd(X, Child, C1, "main").      :- not found_child.
```

To ensure similarity, we add weak constraints (annoted via the dashed gray box in Figure 4). This ensures that a penalty of 1 is added every time we add or delete an individual X to C. Note that the penalty is applied for every context.

The ASP translation of the MR-CKR and the additional constraints constitute the encoding of the problem in ASP. To obtain solutions, we then used a standard ASP solver, namely clingo [16].

Example 5. *Assume our input scene contains four objects* $i_1, \ldots, i_4$ *and* $Child(i_2)$, $Child(i_3)$, $Car(i_4)$ *hold. Due to the ontology axiom* $Child \sqsubseteq Human$, *we could derive* $Human(i_2)$ *and* $Human(i_3)$.

Thus, in context C_2, *where we need a rolling container but no human, we need to remove* i_2 *and* i_3 *from the* $Child$ *concept and add an object to the* $RollingContainer$ *concept. Thus, a potential modification (restricted to context* C_3*) is represented by the interpretation*

$$\mathcal{I} = \{\text{instd}(i_1, ADD_{RollingContainer}, C_3, \text{“}main\text{”}), \text{instd}(i_2, DEL_{Child}, C_3\text{“}main\text{”}),\\ \text{instd}(i_3, DEL_{Child}, C_3, \text{“}main\text{”})\}.$$

As discussed in the previous example, it has cost $\text{cost}(Add) + 2 \cdot \text{cost}(Del)$, *which is* 3 *since we assign cost* 1 *to addition and deletion respectively.*

As there is no modification of a lower cost, one of the possible generated scenes for C_3 *consists of* $RollingContainer(i_1), Car(i_4)$, *i.e., it contains a rolling container* i_1, *no humans, but a car* i_4.

On the other hand, in the context C_1 *we do not need to perform any modifications, since the original scene already contains a child.*

5.1. Specialized Translation

The original MR-CKR to ASP translation is capable of handling highly complex relations between contexts and supports arbitrary defaults and flexible ontological background knowledge. However, this comes at the cost of an encoding in ASP that is not suitable for our purposes, using large scene graphs with many contexts and concepts.

To circumvent this, we specialized the encoding to our setting. Recall that the (defeasible) axioms of the MR-CKR (see Figure 4, annoted in green) tell us that we guess *either* the addition *or* the non-addition of any named individual X to C, as long as there is no other reason in the ontology that prevents both. As our base ontology is consistent, there can never be a reason in our ontology that prevents the non-addition. Hence the either-or really holds in our setting.

This allows us to use the following rules to encode the (defeasible) axioms above:

```
instd(X,ADD_C,Con,"main") :- instd(X,"Named",Con,"main"),
                              not instd(X,NOADD_C,Con,"main").
instd(X,NOADD_C,Con,"main") :- instd(X,"Named",Con,"main"),
                              not instd(X,ADD_C,Con,"main").
```

Clearly, the same can be done for deletion and non-deletion.

This specialized translation for our setting leads to a significant performance improvement. While the original encoding only allows us to generate new scenes using tiny starting scenes, the improved strategy allows inference of the real world scenes from the ontology within seconds.

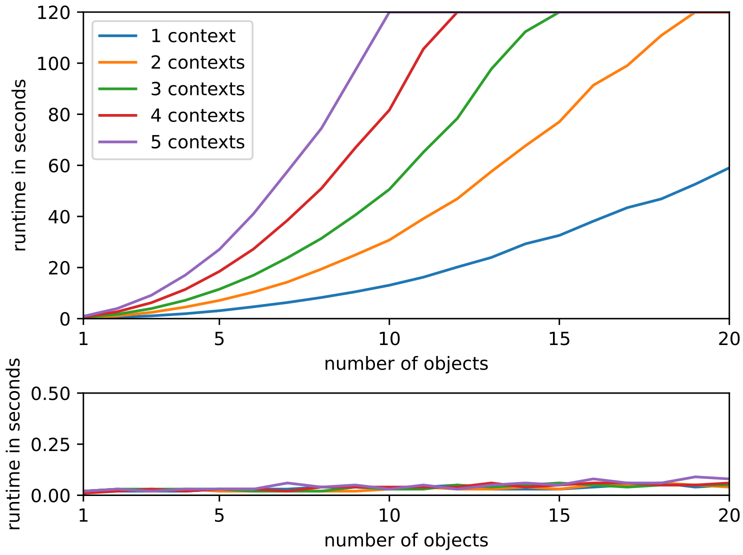

Figure 5: Solving time after using the General (top) and Specialized (bottom) translations.

5.2. Scalability

We briefly investigate how large the instances that we can solve can become, while maintaining a low runtime. Here, we compare the original translation, denoted General, and the specialized one, denoted Specialized.

Secondly, we investigate how much the solving time depends on (i) the number of objects in the scene and (ii) the number of contexts. We vary (i) between 1 and 20 and (ii) between 1 and 5 on a randomly chosen example scene from the ontology.

Here, we use clingo [16] on the generated programs with input option "-t 3" to specify that three threads in parallel should be used to solve the problem. We apply a time limit of 120 seconds and assign runs that do not finish during this time a runtime of 120 seconds.

The results of our investigation are given in Figure 5. We see that even if only one context is used, the runtime after General grows quickly. While it still remains in a feasible range when using one context and up to 20 objects in the scene, the same cannot be said when more contexts are used. For five contexts, solving already becomes slow when ten or more objects are included in the scene. Additionally, the original scene has many more objects (more than 300), thus, this translation can only be employed to restricted examples, even if there is only one context.

On the other hand, for Specialized we see that the solving time is consistently far below one second, even when using all five contexts and 20 objects in the scene. Note here the different limits of the Y-axis, which we adapted to make the runtimes visible. Even using the full scene, which constitutes a realistic industry size example (including more than 300 objects) the solving time remains at around 0.67 seconds. We observed a similar effect in the grounding size, which went to around 80 thousand rules and did not finish with more than half a million rules already produced for the specialized and general encoding, respectively.

We see that while the original translation General is able to handle a broader range of MR-CKRs, it pays off to use the specialized translation Specialized in our setting. With Specialized we can generate new scenes in subsecond times, even if the full scene (i.e., all its objects) and all contexts are used. This suggests that with Specialized we can also generate new inputs for more complex semantic conditions and base ontologies than the ones provided in our prototype, giving us interesting opportunities to extend our work in the future.

6. Conclusion

We have introduced a new framework to generate new interesting inputs for neural models based on existing ones, in particular the setting of scene generation for AV scene data. Notably, our framework does so based on symbolic reasoning methods. This allows us, on the one hand, to incorporate real world knowledge (in the form of contextual knowledge) that ensures that the generated inputs are *realistic*, and, on the other hand, to formulate a semantic criterion that should be satisfied by the new input. We saw that all components that we incorporated in our framework add their respective benefits:

- **MR-CKR** allows us (i) to incorporate ontological knowledge easily and (ii) to perform different modifications in different contexts.
- **Algebraic Measures** allow us to easily specify a cost value to optimize.
- **ASP**, as a declarative backend programming language, allows us to perform reasoning/scene generation efficiently using standard solvers.

While we only considered a small example in our prototype, it successfully generates new scene descriptions. Furthermore, as it can be easily generalized to the generation of different types of scenes, it provides a proof of concept of our approach.

In future work, it will be interesting to extend this example with more complicated semantic descriptions of interesting scenes gathered by inspecting poor performing inputs for a prediction task with a neural model: in particular, it would be interesting to use more complex contextual structures to represent different variations of the scenes, but also use inputs performances to give a quantification of the more interesting cases to be generated. Another open challenge is to use the symbolic description of the new scene to generate images that can be fed to the neural model and assess how much training with these new examples improves the network performance.

Acknowledgments. We would like to thank Michael Pfeiffer, Nicole Finnie, Grace Hua and Jan-Hendrik Metzen from Bosch Center for AI for interesting discussions on the topic of scene generation. This work was partially supported by the European Commission funded projects "Humane AI: Toward AI Systems That Augment and Empower Humans by Understanding Us, our Society and the World Around Us" (grant #820437) and "AI4EU: A European AI on Demand Platform and Ecosystem" (grant #825619), and the Austrian Science Fund (FWF) project W1255-N23. The support is gratefully acknowledged.

References

[1] A. Rozsa, E. M. Rudd, T. E. Boult, Adversarial diversity and hard positive generation, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2016, Las Vegas, NV, USA, June 26 - July 1, 2016, IEEE Computer Society, 2016, pp. 410–417. URL: https://doi.org/10.1109/CVPRW.2016.58. doi:10.1109/CVPRW.2016.58.

[2] B. Yang, H. Zhang, Z. Li, Y. Zhang, K. Xu, J. Wang, Adversarial example generation with adabelief optimizer and crop invariance, Appl. Intell. 53 (2023) 2332–2347. URL: https://doi.org/10.1007/s10489-022-03469-5. doi:10.1007/s10489-022-03469-5.

[3] B. Chen, Y. Feng, T. Dai, J. Bai, Y. Jiang, S. Xia, X. Wang, Adversarial examples generation for deep product quantization networks on image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 45 (2023) 1388–1404. URL: https://doi.org/10.1109/TPAMI.2022.3165024. doi:10.1109/TPAMI.2022.3165024.

[4] J. Gao, J. Lanchantin, M. L. Soffa, Y. Qi, Black-box generation of adversarial text sequences to evade deep learning classifiers, in: 2018 IEEE Security and Privacy Workshops, SP Workshops 2018, San Francisco, CA, USA, May 24, 2018, IEEE Computer Society, 2018, pp. 50–56. URL: https://doi.org/10.1109/SPW.2018.00016. doi:10.1109/SPW.2018.00016.

[5] L. Bozzato, T. Eiter, L. Serafini, Enhancing context knowledge repositories with justifiable exceptions, Artif. Intell. 257 (2018) 72–126.

[6] L. Bozzato, T. Eiter, R. Kiesel, Reasoning on multirelational contextual hierarchies via answer set programming with algebraic measures, Theory Pract. Log. Program. 21 (2021) 593–609. URL: https://doi.org/10.1017/S1471068421000284. doi:10.1017/S1471068421000284.

[7] T. Eiter, R. Kiesel, Weighted LARS for quantitative stream reasoning, in: G. D. Giacomo, A. Catalá, B. Dilkina, M. Milano, S. Barro, A. Bugarín, J. Lang (Eds.), ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020), volume 325 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2020, pp. 729–736. URL: https://doi.org/10.3233/FAIA200160. doi:10.3233/FAIA200160.

[8] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. Patel-Schneider (Eds.), The Description Logic Handbook, Cambridge University Press, 2003.

[9] L. Bozzato, L. Serafini, T. Eiter, Reasoning with justifiable exceptions in contextual hierarchies, in: KR 2018, AAAI Press, 2018, pp. 329–338.

[10] L. Serafini, M. Homola, Contextualized knowledge repositories for the semantic web, J. of Web Semantics 12 (2012) 64–87.

[11] A. Kimmig, G. V. den Broeck, L. D. Raedt, Algebraic model counting, J. Appl. Log. 22 (2017) 46–62. URL: https://doi.org/10.1016/j.jal.2016.11.031. doi:10.1016/j.jal.2016.11.031.

[12] M. Droste, P. Gastin, Weighted automata and weighted logics, Theor. Comput. Sci. 380 (2007) 69–86. URL: https://doi.org/10.1016/j.tcs.2007.02.055. doi:10.1016/j.tcs.2007.02.055.

[13] F. Buccafurri, N. Leone, P. Rullo, Strong and weak constraints in disjunctive datalog, in: J. Dix, U. Furbach, A. Nerode (Eds.), Logic Programming and Nonmonotonic Reasoning, 4th International Conference, LPNMR'97, Dagstuhl Castle, Germany, July 28-31, 1997,

Proceedings, volume 1265 of *Lecture Notes in Computer Science*, Springer, 1997, pp. 2–17. URL: https://doi.org/10.1007/3-540-63255-7_2. doi:10.1007/3-540-63255-7_2.

[14] S. N. Chowdhury, R. Wickramarachchi, M. H. Gad-Elrab, D. Stepanova, C. A. Henson, Towards leveraging commonsense knowledge for autonomous driving., in: ISWC (Posters/Demos/Industry), 2021.

[15] R. Wickramarachchi, C. Henson, A. Sheth, Knowledge-infused learning for entity prediction in driving scenes, Frontiers in big Data (2021) 98.

[16] M. Gebser, R. Kaminski, B. Kaufmann, T. Schaub, Clingo = ASP + control: Preliminary report, CoRR abs/1405.3694 (2014). URL: http://arxiv.org/abs/1405.3694. arXiv:1405.3694.