A Neural-symbolic Framework under Statistical Relational Learning

基于统计关系学习的神经符号框架

Dongran Yu, Xueyan Liu, Shirui Pan, Anchen Li and Bo Yang
于东然，刘雪燕，潘世锐，李安辰，杨波

Abstract-A key objective in the field of artificial intelligence is to develop cognitive models that can exhibit human-like intellectual capabilities. One promising approach to achieving this is through neural-symbolic systems, which combine the strengths of deep learning and symbolic reasoning. However, current methodologies in this area face limitations in integration, generalization, and interpretability. To address these challenges, we propose a neural-symbolic framework based on statistical relational learning, referred to as NSF-SRL. This framework effectively integrates deep learning models with symbolic reasoning in a mutually beneficial manner. In NSF-SRL, the results of symbolic reasoning are utilized to refine and correct the predictions made by deep learning models, while deep learning models enhance the efficiency of the symbolic reasoning process. Through extensive experiments, we demonstrate that our approach achieves high performance and exhibits effective generalization in supervised learning, weakly supervised and zero-shot learning tasks. Furthermore, we introduce a quantitative strategy to evaluate the interpretability of the model's predictions, visualizing the corresponding logic rules that contribute to these predictions and providing insights into the reasoning process. We believe that this approach sets a new standard for neural-symbolic systems and will drive future research in the field of general artificial intelligence.
摘要——人工智能领域的一个关键目标是开发能够展现类人智能能力的认知模型。一种有前景的途径是通过神经符号系统，将深度学习与符号推理的优势结合起来。然而，当前该领域的方法在整合性、泛化能力和可解释性方面存在局限。为解决这些挑战，我们提出了一种基于统计关系学习（Statistical Relational Learning, SRL）的神经符号框架，称为NSF-SRL。该框架有效地实现了深度学习模型与符号推理的互惠整合。在NSF-SRL中，符号推理的结果用于优化和纠正深度学习模型的预测，而深度学习模型则提升了符号推理过程的效率。通过大量实验，我们证明了该方法在监督学习、弱监督学习及零样本学习任务中均表现出高性能和良好的泛化能力。此外，我们引入了一种定量策略来评估模型预测的可解释性，直观展示了促成预测的逻辑规则，揭示了推理过程。我们相信该方法为神经符号系统树立了新标杆，并将推动通用人工智能领域的未来研究。

Index Terms-Neural-symbolic systems, Deep learning, Statistical relational learning, Markov logic networks.
关键词——神经符号系统，深度学习，统计关系学习，马尔可夫逻辑网络。

1 1 INTRODUCTION

2 1 引言

HUMAN cognitive systems encompass both perception and reasoning. Specifically, perception is primarily responsible for recognizing information, while reasoning handles logical deduction and analytical thinking. When humans process information, they integrate both perception and reasoning to enhance their comprehension and decision-making capabilities. Current artificial intelligence systems typically specialize in either perception or reasoning. For instance, deep learning models excel in perception, achieving remarkable performance in tasks that involve inductive learning and computational efficiency. In contrast, symbolic logic is adept at logical reasoning, providing strong results in deductive reasoning tasks, generalization, and interpretability. However, both models have inherent limitations. Deep learning models often operate as black boxes, lacking interpretability, generalizing poorly, and requiring vast amounts of training data to perform optimally. On the other hand, symbolic logic relies on search algorithms to explore solution spaces, resulting in slow reasoning in large-scale environments. Therefore, integrating the strengths of both models offers a way to combine perception and reasoning into a unified framework

that more effectively mimics human cognitive processes. As Leslie G. Valiant argues, reconciling the statistical nature of learning with the logical nature of reasoning to create cognitive computing models that integrate concept learning and manipulation is one of the three fundamental challenges in computer science [1].
人类认知系统涵盖感知与推理两个方面。具体而言，感知主要负责信息识别，而推理则处理逻辑演绎和分析思维。人类在处理信息时，将感知与推理相结合，以增强理解和决策能力。当前的人工智能系统通常专注于感知或推理其中之一。例如，深度学习模型在感知方面表现卓越，在归纳学习和计算效率任务中取得显著成果。相比之下，符号逻辑擅长逻辑推理，在演绎推理、泛化能力和可解释性方面表现优异。然而，两者均存在固有限制。深度学习模型常被视为“黑箱”，缺乏可解释性，泛化能力较差，且需大量训练数据以达到最佳性能。符号逻辑则依赖搜索算法探索解空间，在大规模环境中推理速度较慢。因此，整合两者优势，构建一个统一框架，将感知与推理结合起来，更有效地模拟人类认知过程，成为一种可行路径。正如Leslie G. Valiant所言，将学习的统计性质与推理的逻辑性质相融合，创建集概念学习与操作于一体的认知计算模型，是计算机科学的三大基本挑战之一[1]。

The neural-symbolic system represents a promising approach for effectively integrating perception and reasoning into a unified framework [2], [3], [4]. Various neural-symbolic systems have been proposed, which can be broadly classified into three categories [5]: learning-for-reasoning methods, reasoning-for-learning methods, and learning-reasoning methods. Learning-for-reasoning methods [6], [7], [8] primarily focus on symbolic reasoning. In these methods, deep learning models transform unstructured inputs into symbolic representations, which are then processed by symbolic reasoning models to derive solutions. In some cases, deep learning models replace search algorithms, thus accelerating symbolic reasoning. Reasoning-for-learning approaches [9], [10], [11] focus more on deep learning. Symbolic knowledge is encoded into distributed representations and integrated into deep learning models to compute results. However, these methods often use deep learning to support symbolic reasoning or incorporate symbolic priors to enhance deep learning, without fully achieving complementary integration. Few studies explore learning-reasoning methods, which aim for more comprehensive integration [12], [13]. For example, Manhaeve et al. [14] combine a deep learning model with a probabilistic logic programming language, where the output of the deep learning model serves as input for symbolic reasoning. Techniques like arithmetic circuits and gradient semi-rings enable interaction between the deep learning model and symbolic reasoning. Zhou [13] integrates machine learning with logic reasoning based on the principle of abduction, using the machine learning model's output as input for logical reasoning. This reasoning process iteratively corrects the model's output through consistency optimization, and the refined output is then used as supervised information for further training. While these approaches represent significant progress in neural-symbolic systems, achieving full integration remains a challenging and open problem, necessitating further exploration and research.
神经符号系统是一种将感知与推理有效整合于统一框架的有前景方法[2]，[3]，[4]。现有多种神经符号系统，通常可分为三类[5]：面向推理的学习方法、面向学习的推理方法和学习-推理方法。面向推理的学习方法[6]，[7]，[8]主要聚焦符号推理。在此类方法中，深度学习模型将非结构化输入转化为符号表示，随后由符号推理模型处理以获得解答。在某些情况下，深度学习模型替代搜索算法，从而加速符号推理。面向学习的推理方法[9]，[10]，[11]更侧重深度学习。符号知识被编码为分布式表示并融入深度学习模型以计算结果。然而，这些方法通常仅利用深度学习支持符号推理或引入符号先验以增强深度学习，未能实现完全互补的整合。少数研究探讨学习-推理方法，旨在实现更全面的整合[12]，[13]。例如，Manhaeve等[14]结合深度学习模型与概率逻辑编程语言，深度学习模型的输出作为符号推理的输入。算术电路和梯度半环等技术实现了深度学习模型与符号推理的交互。Zhou[13]基于溯因推理原理，将机器学习与逻辑推理结合，利用机器学习模型输出作为逻辑推理输入。该推理过程通过一致性优化迭代修正模型输出，精炼结果再用作监督信息进行进一步训练。尽管这些方法在神经符号系统领域取得了重要进展，实现完全整合仍是一个具有挑战性且未解决的问题，亟需进一步探索与研究。

- B. Yang (corresponding author), X. Liu (corresponding author) and A. Li are with the Key Laboratory of Symbolic Computation and Knowledge Engineer, Ministry of Education, Jilin University, Changchun, Jilin 130012, China and the School of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China. E-mail: ybo@jlu.edu.cn; xueyanliu@jlu.edu.cn; liac20@mails.jlu.edu.cn

- B. Yang（通讯作者）、X. Liu（通讯作者）和A. Li隶属于吉林大学教育部符号计算与知识工程重点实验室，地址为吉林省长春市吉林大学，邮编130012，同时也在吉林大学计算机科学与技术学院。电子邮箱：ybo@jlu.edu.cn；xueyanliu@jlu.edu.cn；liac20@mails.jlu.edu.cn

- D. Yu is with the Key Laboratory of Symbolic Computation and Knowledge Engineer, Ministry of Education, Jilin University, Changchun, Jilin 130012, China, and the School of Artificial Intelligence, Jilin University, Changchun, Jilin 130012, China.
- D. Yu隶属于吉林大学教育部符号计算与知识工程重点实验室，地址为吉林省长春市吉林大学，邮编130012，同时也在吉林大学人工智能学院，地址同上。

E-mail: yudran@foxmail.com
电子邮箱：yudran@foxmail.com

- S. Pan is with School of Information and Communication Technology, Griffith University, Brisbane 4222, Queensland, Australia. E-mail: s.pan@griffith.edu.au
- S. Pan隶属于澳大利亚昆士兰州布里斯班4222格里菲斯大学信息与通信技术学院。电子邮箱：s.pan@griffith.edu.au

---

This paper introduces a novel framework called the Neural Symbolic Framework under Statistical Relational Learning (NSF-SRL for short), which aims to integrate deep learning models with symbolic logic in a mutually beneficial manner. In NSF-SRL, symbolic logic enhances deep learning models by making their predictions more logical, consistent with common sense, and interpretable, thereby improving their generalization capabilities. In turn, deep learning enhances symbolic logic by increasing its efficiency and robustness to noise. However, a key challenge in constructing the NSF-SRL framework is determining how to effectively combine deep learning and symbolic logic to model a joint probability distribution.
本文提出了一种名为统计关系学习下的神经符号框架（Neural Symbolic Framework under Statistical Relational Learning，简称NSF-SRL）的新型框架，旨在实现深度学习模型与符号逻辑的互利融合。在NSF-SRL中，符号逻辑通过使深度学习模型的预测更具逻辑性、符合常识且可解释，从而提升其泛化能力；反过来，深度学习则通过提高符号逻辑的效率和抗噪声能力来增强其表现。然而，构建NSF-SRL框架的关键挑战在于如何有效结合深度学习与符号逻辑以建模联合概率分布。

Statistical Relational Learning (SRL) [15] serves as a bridge between statistical models, such as deep learning, and relational models, like symbolic logic, by integrating the two approaches. Inspired by this framework, we employs SRL techniques to address the challenge of model construction. In this approach, deep learning processes data according to specific tasks and generates corresponding outputs, while symbolic logic learns a joint probability distribution based on these outputs and symbolic knowledge, thus constraining deep learning's predictions to achieve mutual enhancement. It is important to note that in our framework, deep learning not only functions as a data processor for symbolic logic but also replaces traditional search algorithms to improve computational efficiency. In this study, symbolic knowledge is represented using First-Order Logic (FOL). During the training phase,the model learns the basic concepts 1 in FOL from the sample data, a process we term concept learning. In the testing phase, the model utilizes existing or newly acquired FOLs to combine and manipulate learned concepts, thereby generating new ones-a process referred to as concept manipulation.
统计关系学习（Statistical Relational Learning，SRL）[15]作为统计模型（如深度学习）与关系模型（如符号逻辑）之间的桥梁，将两者融合。受此框架启发，我们采用SRL技术解决模型构建难题。在该方法中，深度学习根据特定任务处理数据并生成相应输出，而符号逻辑基于这些输出及符号知识学习联合概率分布，从而约束深度学习的预测，实现互相促进。值得注意的是，在我们的框架中，深度学习不仅作为符号逻辑的数据处理器，还取代传统搜索算法以提升计算效率。本研究中，符号知识采用一阶逻辑（First-Order Logic，FOL）表示。训练阶段，模型从样本数据中学

习FOL中的基本概念，此过程称为概念学习。测试阶段，模型利用已有或新获得的FOL结合并操作已学概念，生成新概念，此过程称为概念操作。

In summary, our contributions can be characterized in threefold:
综上所述，我们的贡献可归纳为三点：

- In this study, we propose a general neural-symbolic system framework NSF-SRL and develop an end-to-end model.
- 本研究提出了通用的神经符号系统框架NSF-SRL，并开发了端到端模型。

- The model employs statistical relational learning techniques to integrate deep learning and symbolic logic, thereby achieving mutual enhancement of learning and reasoning. This integration improves the model's generalization ability and interpretability.
- 该模型采用统计关系学习技术融合深度学习与符号逻辑，实现学习与推理的相互促进，提升模型的泛化能力和可解释性。

- Based on our experimental results, we demonstrate that NSF-SRL outperforms comparable methods in various reasoning tasks, including supervised, weakly supervised, and zero-shot learning scenarios, with respect to performance and generalization. Additionally, we emphasize the interpretability of our model by providing visualizations that enhance the understanding of the reasoning process.
- 基于实验结果，我们证明NSF-SRL在多种推理任务中（包括监督学习、弱监督学习和零样本学习场景）在性能和泛化能力上优于同类方法。此外，我们通过可视化展示强调了模型的可解释性，增强了对推理过程的理解。

In our previous conference paper [16], we initially presented and validated the proposed approach for visual relationship detection. However, this current study significantly extends that work by introducing new model designs, such as concept manipulation, incorporating new tasks like digit image addition and zero-shot image classification, and comparing against additional baseline approaches. Furthermore, we provide extensive experimental validations and comparisons to thoroughly evaluate the model's performance.
在我们之前的会议论文[16]中，首次提出并验证了该方法在视觉关系检测中的应用。然而，本研究在此基础上进行了显著扩展，新增了如概念操作等模型设计，涵盖了数字图像加法和零样本图像分类等新任务，并与更多基线方法进行了比较。此外，我们提供了详尽的实验验证和对比，以全面评估模型性能。

3 2 RELATED WORK

4 2 相关工作

Neural-symbolic systems. In recent times, neural-symbolic reasoning has gained significant attention and can be classified into three main groups [5]. The first group consists of methods where deep neural networks assist symbolic reasoning. These methods replace traditional search algorithms in symbolic reasoning with deep neural networks to reduce the search space and improve computation speed [6], [7], [8], [17]. For example, Qu et al. [6] proposed probabilistic Logic Neural Networks (pLogicNet), which addresses the problem of reasoning in knowledge graphs (triplet completion) as an inference problem involving hidden variables in a probabilistic graph model. The pLogicNet employs a combination of variational EM and neural networks to approximate the inference. Building on the idea of pLogicNet, Zhang et al. [7] introduced ExpressGNN, which leverages Graph Neural Networks (GNNs) as approximate inference methods for posterior calculation in the variable EM algorithm. Marra et al. [18] proposed NMLN, which reparametrizes the MLN through a neural network that is evaluated based on input features. The second group focuses on symbolic reasoning aiding deep learning models during the learning process. These methods incorporate symbolic knowledge into the training of deep learning models to enhance performance and interpretability [10], [11], [19], [20], [21]. Symbolic knowledge is often used as a regularizer during training. For instance, Xie et al. [10] encode symbolic knowledge into neural networks by designing a regularization term in the loss function for a specific task. The third group consists of models that strike a balance between deep learning

models and symbolic reasoning, allowing both paradigms to contribute to problem-solving [12], [13], [22], [23]. Zhou [13] establishes a connection between machine learning and symbolic reasoning frameworks based on the characteristics of symbolic reasoning, such as abduction. Duan et al. [24] proposed a framework for joint learning of neural perception and logical reasoning, where the two components are mutually supervised and jointly optimized. Pryor et al. [25] introduced NeuPSL, where the neural network learns the predicates for logical reasoning, while logical reasoning imposes constraints on the neural network. Yang et al., [26] proposed NeurASP, which leverages a pre-trained neural network in symbolic computation and enhances the neural network's performance by applying symbolic reasoning. In contrast to the aforementioned methods, our approach takes a different route to bridge the gap between deep learning models and symbolic logic through statistical relational learning. By leveraging statistical relational learning, our method retains the full capabilities of both probabilistic reasoning and deep learning, offering a unique and powerful integration of the two paradigms.
神经符号系统。近年来，神经符号推理受到了广泛关注，可分为三大类[5]。第一类方法是深度神经网络辅助符号推理。这些方法用深度神经网络替代符号推理中的传统搜索算法，以减少搜索空间并提升计算速度[6]，[7]，[8]，[17]。例如，Qu等人[6]提出了概率逻辑神经网络（pLogicNet），将知识图谱推理（三元组补全）问题视为涉及隐变量的概率图模型中的推断问题。pLogicNet结合变分EM和神经网络来近似推断。基于pLogicNet的思想，Zhang等人[7]提出了ExpressGNN，利用图神经网络（GNN）作为变量EM算法中后验计算的近似推断方法。Marra等人[18]提出了NMLN，通过一个基于输入特征评估的神经网络对马尔可夫逻辑网络（MLN）进行重新参数化。第二类方法关注符号推理在深度学习模型训练过程中的辅助作用，这些方法将符号知识融入深度学习模型的训练中，以提升性能和可解释性[10]，[11]，[19]，[20]，[21]。符号知识通常作为训练过程中的正则项。例如，Xie等人[10]通过设计特定任务的损失函数正则项，将符号知识编码进神经网络。第三类方法在深度学习模型和符号推理之间寻求平衡，使两者共同参与问题解决[12]，[13]，[22]，[23]。Zhou[13]基于符号推理的特性（如溯因推理）建立了机器学习与符号推理框架的联系。Duan等人[24]提出了神经感知与逻辑推理联合学习框架，两者相互监督并联合优化。Pryor等人[25]提出了NeuPSL，神经网络学习逻辑推理的谓词，逻辑推理则对神经网络施加约束。Yang等人[26]提出了NeurASP，利用预训练神经网络进行符号计算，并通过符号推理提升神经网络性能。与上述方法不同，我们的方法通过统计关系学习架起深度学习模型与符号逻辑之间的桥梁。借助统计关系学习，我们的方法保留了概率推理和深度学习的全部能力，实现了两种范式的独特且强大的融合。

Markov Logic Networks. To handle complexity and uncertainty of the real world, intelligent systems require a unified representation that combines first-order logic (FOL) and probabilistic graphical models. Markov Logic Networks (MLNs) achieve this by providing a unified framework that combines FOL and probabilistic graphical models into a single representation. MLN has been extensively studied and proven effective in various reasoning tasks, including knowledge graph reasoning [6], [7], semantic parsing [27], [28], and social network analysis [29]. MLN is capable of capturing complexity and uncertainty inherent in relational data. However, performing inference and learning in MLN can be computationally expensive due to the exponential cost of constructing the ground MLN and NP-complete optimization problem. This limitation hinders the practical application of MLN in large-scale scenarios. To address these challenges, many works have been proposed to improve accuracy and efficiency of MLN. For instance, some studies [30], [31] have focused on enhancing the accuracy of MLN, while others [6], [7], [32], [33], [34] have aimed to improve its efficiency. In particular, two studies [6], [7] have replaced traditional inference algorithms in MLN with neural networks. By leveraging neural networks, these approaches offer a more efficient alternative for performing inference in MLN. This integration of neural networks and MLN allows for more scalable and effective reasoning in large-scale applications.
马尔可夫逻辑网络。为应对现实世界的复杂性和不确定性，智能系统需要一种将一阶逻辑（FOL）与概率图模型结合的统一表示。马尔可夫逻辑网络（MLN）通过提供将FOL与概率图模型融合为单一表示的统一框架，实现了这一目标。MLN已被广泛研究，并在多种推理任务中证明了其有效性，包括知识图谱推理[6]，[7]、语义解析[27]，[28]和社交网络分析[29]。MLN能够捕捉关系数据中固有的复杂性和不确定性。然而，由于构建底层MLN的指数级成本及NP完全的优化问题，MLN的推断和学习计算开销较大，限制了其在大规模场景中的实际应用。为解决这些挑战，许多工作致力于提升MLN的准确性和效率。例如，一些研究[30]，[31]专注于提高MLN的准确性，而另一些[6]，[7]，[32]，[33]，[34]则致力于提升其效率。特别地，两项研究[6]，[7]用神经网络替代了MLN中的传统推断算法。通过利用神经网络，这些方法为MLN推断提供了更高效的替代方案。这种神经网络与MLN的结合，使得在大规模应用中实现更具扩展性和有效性的推理成为可能。

1. In this paper, concepts refer to predicates in FOL.
2. 本文中，概念指一阶逻辑（FOL）中的谓词。

5 3 PRELIMINARIES

6 3 预备知识

In this section, we first introduce the neural-symbolic model definition and notations in this paper. Then, we will introduce the basic knowledge about statistic relational learning.
本节首先介绍本文中的神经符号模型定义和符号表示，随后介绍统计关系学习的基础知识。

6.1 3.1 Model Description

6.2 3.1 模型描述

The primary task in developing the model NSF-SRL is to formulate and maximize the posterior probability $P(Y \mid X, R; \theta_1, \theta_2, w)$, where $X = \{x_1, x_2, \ldots, x_n\}$ represents the observed data, $Y = \{y_1, y_2, \ldots, y_n\}$ is the label set corresponding to data X ,and $R = \{r_1, r_2, \ldots, r_m\}$ is the first-order logic rule set, θ_1, θ_2 and w denote the parameters of NSF-SRL. n is the number of the instance of raw data,and m is the number of rules. Given the training dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ and the first-order logic rules R ,the learning process of NSF-SRL can be expressed as maximizing the posterior probability, formally defined as:
开发NSF-SRL模型的主要任务是构建并最大化后验概率$P(Y \mid X, R; \theta_1, \theta_2, w)$，其中$X = \{x_1, x_2, \ldots, x_n\}$表示观测数据，$Y = \{y_1, y_2, \ldots, y_n\}$是对应数据$X$的标签集，$R = \{r_1, r_2, \ldots, r_m\}$是一阶逻辑规则集，$\theta_1, \theta_2$和$w$表示NSF-SRL的参数。$n$是原始数据实例的数量，$m$是规则的数量。给定训练数据集$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$和一阶逻辑规则$R$，NSF-SRL的学习过程可以表示为最大化后验概率，形式定义如下：

$$\forall D \max_{\theta_1, \theta_2, w} P(Y \mid X, R; \theta_1, \theta_2, w), \tag{1}$$

For example,in image classification tasks,the input data D represents images,while the output y corresponds to the labels of the objects within those images. To enhance understanding of this paper, symbolic descriptions are provided in Table 1 These descriptions clarify the symbolic representations used throughout the study and facilitate comprehension of the concepts and methodologies discussed.
例如，在图像分类任务中，输入数据D表示图像，而输出y对应图像中物体的标签。为增强对本文的理解，表1提供了符号描述，这些描述阐明了研究中使用的符号表示，有助于理解所讨论的概念和方法。

6.3 3.2 Statistical Relational Learning

6.4 3.2 统计关系学习

Many tasks in real-world application domains are characterized by the presence of both uncertainty and complex relational structures. Statistical learning addresses the former, while relational learning focuses on the latter. Statistical Relational Learning (SRL) aims to harness the strengths of both approaches [15].
现实应用领域中的许多任务同时具有不确定性和复杂的关系结构。统计学习解决前者，关系学习关注后者。统计关系学习（Statistical Relational Learning，SRL）旨在结合两者的优势[15]。

In this study, we leverage SRL to integrate first-order logic (FOL,rule body ⇒ rule head) with probabilistic graphical models, creating a unified framework that facilitates probabilistic inference for reasoning problems. FOL represents a type of commonsense (symbolic) knowledge that is easily understood by humans. In this paper, we treat the FOL language as a means to describe knowledge in the form of logic rules, which provides strong expressive capability [35]. For instance, FOL allows for the definition of predicates and the description of various relations.
本研究利用SRL将一阶逻辑（FOL，规则体⇒规则头）与概率图模型相结合，构建统一框架以便于推理问题的概率推断。FOL代表一种人类易于理解的常识（符号）知识。本文将FOL语言视为描述逻辑规则形式知识的工具，具备强大的表达能力[35]。例如，FOL允许定义谓词并描述各种关系。

To achieve this integration, we employ Markov Logic Networks (MLNs), a well-known statistical relational learning model, to represent FOL as undirected graphs. In the constructed undirected graph, nodes are generated based on all ground atoms which are logical predicates with their arguments replaced by specific constants. In this paper, a_r denotes assignments of variables to the arguments of an FOL r ,and all consistent assignments are captured in the set $A_r = \left\{a_r^1, a_r^2, \ldots\right\}$. For instance, if we have a constant set $C = \{c_1, c_2\}$ and an FOL $r \in R$ such as catlike $(x) \wedge \text{tawn}\, y\,(x) \wedge \text{spot}\,(x) \Rightarrow$ leopard(x), the corresponding ground atoms A_r can be generated such as {catlike (c_1) ,catlike (c_2) ,tawny (c_1) ,tawny (c_2) ,spot (c_1) ,spot (c_2) ,leopard $(c_1), \text{leopard}\,(c_2)\}$. Furthermore,an edge is established between two nodes if the corresponding ground atoms co-occur in at least one ground FOL in the MLN. Consequently, a ground MLN can be formulated as a joint probability distribution, capturing the dependencies and correlations among the ground atoms. This joint probability distribution is expressed as:
为实现该整合，我们采用著名的统计关系学习模型马尔可夫逻辑网络（Markov Logic Networks，MLNs）将FOL表示为无向图。在构建的无向图中，节点基于所有基元原子生成，基元原子是将逻辑谓词的参数替换为具体常量后的表达。本文中，a_r表示变量对FOLr参数的赋值，所有一致的赋值被包含在集合$A_r = \left\{a_r^1, a_r^2, \ldots\right\}$中。例如，若有常量集合$C = \{c_1, c_2\}$和FOL$r \in R$如catlike $(x) \wedge \text{tawn}\, y\,(x) \wedge \text{spot}\,(x) \Rightarrow$ leopard(x)，则可生成相应的基元原子A_r，如{catlike (c_1)，catlike (c_2)，tawny (c_1)，tawny (c_2)，spot (c_1)，spot (c_2)，leopard $(c_1), \text{leopard}\,(c_2)\}$}。此外，若两个基元原子在MLN中的至少一个基元FOL中共现，则在对应节点间建立边。因此，基元MLN可被表述为联合概率分布，捕捉基元原子间的依赖和关联。该联合概率分布表达为：

$$P\,(A) = \frac{1}{Z\,(w)} \exp\left\{\sum_{r \in R} w_r \sum_{a_r \in A_r} \phi\,(a_r)\right\}, \tag{2}$$

TABLE 1

Important notations and their descriptions.
重要符号及其说明。

Notations	Descriptions
D	Set of input data
Y	Set of ground truths
$\widehat{y}$	Pseudo-label
R	Set of logical rules
r	A logic rule
${T}_{r}$	Triggered logic rule
A	Ground atom sets in knowledge base
${A}_{r}$	Ground atom sets in a logic rule
${a}_{r}$	A ground atom
ϕ	Potential function
${\theta }_{1}$	Parameters of neural reasoning module
${\theta }_{2}$	Parameters of concept network
W	Weight sets of the logic rules
${w}_{r}$	Weight of a logic rule

符号表示	描述
D	输入数据集
Y	真实标签集
$\widehat{y}$	伪标签
R	逻辑规则集
r	一条逻辑规则
${T}_{r}$	触发的逻辑规则
A	知识库中的基元原子集
${A}_{r}$	逻辑规则中的基元原子集
${a}_{r}$	一个基元原子
ϕ	势函数
${\theta }_{1}$	神经推理模块的参数
${\theta }_{2}$	概念网络的参数
W	逻辑规则的权重集
${w}_{r}$	一条逻辑规则的权重

where $Z(w) = \sum_{A} \sum_{A_r \in A, a_r \in A_r} \phi(a_r)$ is the partition function that sums over all ground atoms. A represents all ground atoms in the knowledge base,while ϕ is a potential function reflecting the number of times a FOL statement is true. The variable w denotes the weight sets of all FOLs,and w_r refers to the weight of a specific FOL.
其中 $Z(w) = \sum_{A} \sum_{A_r \in A, a_r \in A_r} \phi(a_r)$ 是对所有基原子求和的配分函数。A 表示知识库中的所有基原子，而 ϕ 是反映一阶逻辑（FOL）语句为真的次数的势函数。变量 w 表示所有一阶逻辑的权重集合，w_r 指特定一阶逻辑的权重。

7 4 OUR METHOD: NSF-SRL

8 4 我们的方法：NSF-SRL

The goal of the NSF-SRL framework is to achieve a mutual integration of deep learning and symbolic logic. In this framework, deep learning can take the form of any task-related neural network, primarily responsible for feature extraction and result prediction. Symbolic logic, on the other hand, is grounded in probabilistic graphical models and is responsible for logical reasoning. In this section, we first provide an overview of our NSF-SRL in Section 4.1.

We then present concept learning in Section 4.2, followed by a description of concept manipulation in Section 4.3
NSF-SRL框架的目标是实现深度学习与符号逻辑的相互融合。在该框架中，深度学习可以采用任何与任务相关的神经网络形式，主要负责特征提取和结果预测。符号逻辑则基于概率图模型，负责逻辑推理。本节中，我们首先在4.1节概述NSF-SRL，随后在4.2节介绍概念学习，最后在4.3节描述概念操作。

---

2. Ground atom is a replacement of all of its arguments by constants. In this paper, we refer to the process of replacement as "grounding".
3. 基原子是将其所有参数替换为常量的结果。本文中，我们将该替换过程称为“基化”（grounding）。

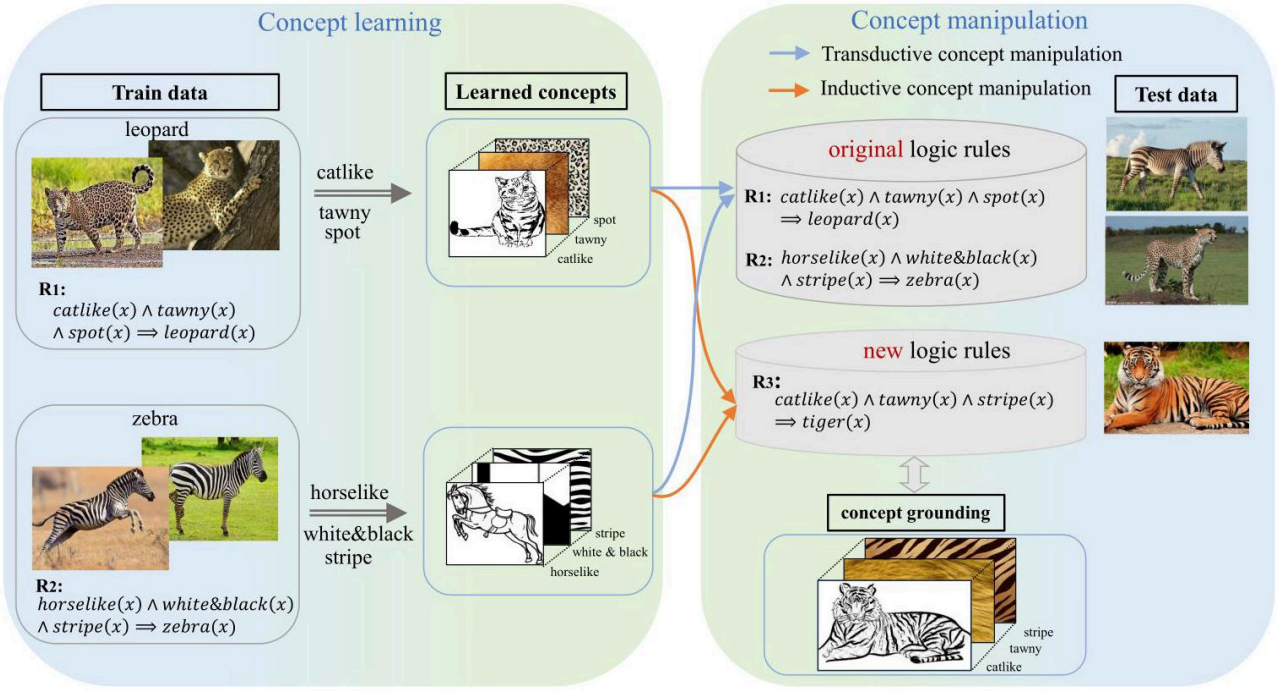


Fig. 1. Overview of NSF-SRL. The concept learning phase acquires basic concepts such as "cattike", "tawny" and "spot" from the training data. In transductive concept manipulation, the learned concepts and toriginal rules are applied to test data whose labels were present in the training sets. This integration of learned concepts enhances the interpretability of NSF-SRL by providing insights into how predictions are made based on these concepts and the accompanying rules. Conversely, in inductive concept manipulation, the learned concepts serve as the rule body, and new rules are introduced to reason about samples with labels that have never appeared in the training set.
图1. NSF-SRL概览。概念学习阶段从训练数据中获取基本概念，如“猫状”（catlike）、“黄褐色”（tawny）和“斑点”（spot）。在传导式概念操作中，学习到的概念和原始规则被应用于训练集中已出现标签的测试数据。通过结合学习到的概念，NSF-SRL的可解释性得以增强，揭示了基于这些概念及相应规则如何做出预测。相反，在归纳式概念操作中，学习到的概念作为规则体，引入新规则对训练集中未出现标签的样本进行推理。

8.1 4.1 Overview of NSF-SRL

8.2 4.1 NSF-SRL概览

An overview of the NSF-SRL framework, consisting of two key phases-concept learning and concept manipulation-is presented in Fig. 1
NSF-SRL框架概览，包括两个关键阶段——概念学习和概念操作，如图1所示。

Concept learning focuses on acquiring fundamental concepts from training data. For instance, we can learn essential concepts such as "catlike", "tawny" and "spot" from images of leopards, and "horselike", "white&black" and "stripe" from images of zebras,utilizing the rules $\mathbf{R1}: \mathbf{catlike}(x) \wedge \mathbf{tawny}(x) \wedge \mathbf{spot}(x) \Rightarrow$ leopard(x)and R2: horselike $(x) \wedge$ white&black $(x) \wedge$ stripe $(x) \Rightarrow \mathbf{zebra}(x)$.
概念学习侧重于从训练数据中获取基本概念。例如，我们可以从豹的图像中学习“猫状”（catlike）、“黄褐色”（tawny）和“斑点”（spot）等基本概念，从斑马的图像中学习“马状”（horselike）、“黑白相间”（white&black）和“条纹”（stripe）等概念，利用规则 $\mathbf{R1}: \mathbf{catlike}(x) \wedge \mathbf{tawny}(x) \wedge \mathbf{spot}(x) \Rightarrow$ leopard(x) 和 R2: horselike $(x) \wedge$ white&black $(x) \wedge$ stripe $(x) \Rightarrow \mathbf{zebra}(x)$ 。

Concept manipulation is used for reasoning and interpreting results, employing existing or newly acquired symbolic knowledge to combine established concepts and generate new ones. In this paper, we identify two types of conceptual operations: transduc-tive concept manipulation and inductive concept manipulation. In transductive concept manipulation, the learned concepts and original rules are utilized to test data whose labels have appeared in the training set. Incorporating these learned concepts enhances the interpretability of the NSF-SRL, providing insights into how prediction results are derived in conjunction with the rules. For example, the predicted label "leopard" can be attributed to rule R1. Conversely, in inductive concept manipulation, the learned concepts and the new rules are applied to test data whose label has never appeared in the training set. Specifically, the learned concepts serve as the rule body of a new rule, which is used to reason the rule head as the output when testing a new sample. For instance, when an image containing a tiger is fed into the well-trained model, it can trigger the new rule R3 and generate corresponding ground atoms such as "catlike", "tawny" and "stripe" via concept grounding. By leveraging R3 and the ground atoms, the model infers the new concept "tiger". Inductive concept manipulation enables the application of previously learned concepts to new tasks, facilitating the generation of new concepts through inference and realizing adaptation and generalization to new tasks. In summary, through the process of concept manipulation, the NSF-SRL effectively learns, reasons, and produces explainable results by leveraging learned concepts.
概念操作用于推理和解释结果，利用已有或新获得的符号知识结合已建立的概念生成新概念。本文中，我们识别出两种概念操作类型：传导式概念操作和归纳式概念操作。在传导式概念操作中，学习到的概念和原始规则被用于训练集中已出现标签的测试数据。引入这些学习到的概念提升了NSF-SRL的可解释性，揭示了预测结果如何结合规则得出。例如，预测标签“豹”可归因于规则R1。相反，在归纳式概念操作中，学习到的概念和新规则被应用于训练集中未出现标签的测试数据。具体而言，学习到的概念作为新规则的规则体，用于推理规则头作为测试新样本时的输出。例如，当一张包含老虎的图像输入训练良好的模型时，可触发新规则R3，并通过概念基化生成“猫状”、“黄褐色”和“条纹”等基原子。借助R3和这些基原子，模型推断出新概念“老虎”。归纳式概念操作使得先前学习的概念能够应用于新任务，通过推理生成新概念，实现对新任务的适应和泛化。总之，通过概念操作过程，NSF-SRL有效地利用学习到的概念进行学习、推理并产生可解释的结果。

8.3 4.2 Concept Learning

8.4 4.2 概念学习

Concept learning involves a Neural Reasoning Module (NRM) and a Symbolic Reasoning Module (SRM), as illustrated in Fig. 2 These two modules engage in end-to-end joint learning to produce a trained model. Specifically, the NRM functions as a task network, generating pseudo-labels and feature vectors. In contrast, the SRM operates as a probabilistic graphical model responsible for deriving reasoning outcomes. During the training process, the SRM constrains the parameter learning of the NRM, enhancing the accuracy and interpretability of its predictions.

After N iterations and corresponding parameter updates,the trained model is achieved.
概念学习包括神经推理模块（Neural Reasoning Module, NRM）和符号推理模块（Symbolic Reasoning Module, SRM），如图2所示。这两个模块通过端到端的联合学习生成训练模型。具体而言，NRM作为任务网络，生成伪标签和特征向量；而SRM作为概率图模型，负责推导推理结果。在训练过程中，SRM约束NRM的参数学习，提高其预测的准确性和可解释性。经过N次迭代及相应的参数更新后，获得训练好的模型。

8.4.1 4.2.1 Neural Reasoning Module

8.4.2 4.2.1 神经推理模块

The Neural Reasoning Module (NRM) is a versatile deep neural network whose architecture can vary according to the specific task at hand. This adaptability enables the NRM to accommodate diverse tasks and to be implemented with various network architectures. For instance, in the digital image addition task, the NRM may utilize a Convolutional Neural Network (CNN) to process image data, whereas in object detection, it may adopt a network structure incorporating ResNet to enhance detection performance. This capability to dynamically adjust the network architecture based on task requirements allows the NRM to effectively meet the needs of different applications. The objective to be maximized in terms of log-likelihood is formalized as follows:
神经推理模块（NRM）是一种多功能深度神经网络，其架构可根据具体任务灵活调整。这种适应性使NRM能够处理多样化任务，并采用不同的网络结构实现。例如，在数字图像加法任务中，NRM可能使用卷积神经网络（Convolutional Neural Network, CNN）处理图像数据；而在目标检测中，则可能采用包含残差网络（ResNet）的结构以提升检测性能。基于任务需求动态调整网络架构的能力，使NRM能够有效满足不同应用的需求。最大化的对数似然目标形式化如下：

$$O_{\text{task}} = \log P_{\theta_1}(\hat{y} \mid D), \tag{3}$$

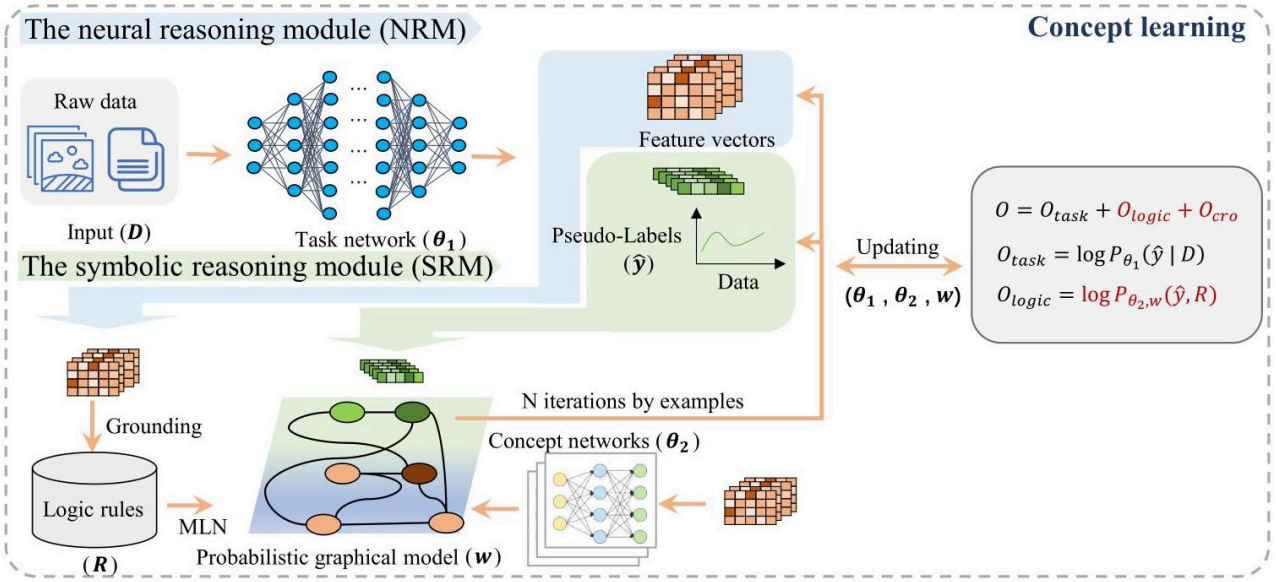


Fig. 2. Illustration of concept learning. The NRM aims to predict labels for raw data, generating pseudo-labels and feature vectors as outputs. The SRM is a probabilistic graphical model that incorporates both the pseudo-labels from the NRM and the ground atoms from the MLN. The entire model is trained end-to-end, using backpropagation to iteratively refine the pseudo-labels.
图2. 概念学习示意图。NRM旨在为原始数据预测标签，输出伪标签和特征向量。SRM是一个概率图模型，结合了

NRM的伪标签和马尔可夫逻辑网络（Markov Logic Network, MLN）中的基础原子。整个模型通过反向传播端到端训练，迭代优化伪标签。

where θ_1 is the learnable parameter of the NRM. At the beginning of the model training, the NRM may produce predictions with substantial errors due to insufficient training. Consequently, in this paper,we refer to these predictions as pseudo-labels $\hat{y}$.
其中θ_1是NRM的可学习参数。在模型训练初期，由于训练不足，NRM可能产生较大误差的预测。因此，本文将这些预测称为伪标签$\hat{y}$。

8.4.3 4.2.2 Symbolic Reasoning Module

8.4.4 4.2.2 符号推理模块

The Symbolic Reasoning Module (SRM) plays a critical role in supporting the NRM by facilitating learning and employing reasoning to generate predictive outcomes and provide evidence for result interpretation. Specifically, the SRM operates as follows: when presented with a training sample (x_i, y_i) ,it is responsible for deducing the outcome y_i based on the predicted label $\hat{y}_i$,the feature vector output by the NRM, and first-order logic rules. If $\widehat{y_i}$ is incorrect,the SRM adjusts the NRM parameters through backpropagation to correct the prediction. To achieve this, we leverage SRL to construct a probabilistic graphical model within the SRM, as depicted in Fig. 2. The primary objective of the SRM is to utilize SRLs for learning variables and guiding the NRM's reasoning in the correct direction, effectively serving as an error corrector. In this study, the probabilistic graphical model is instantiated using a MLN that encompasses all tasks discussed in the validations.
符号推理模块（SRM）在支持NRM方面发挥关键作用，通过促进学习和推理生成预测结果，并为结果解释提供证据。具体而言，SRM的运作如下：给定训练样本(x_i, y_i)，SRM基于预测标签$\hat{y}_i$、NRM输出的特征向量及一阶逻辑规则推导结果y_i。若$\widehat{y_i}$不正确，SRM通过反向传播调整NRM参数以纠正预测。为此，我们利用统计关系学习（Statistical Relational Learning, SRL）构建SRM中的概率图模型，如图2所示。SRM的主要目标是利用SRL学习变量并引导NRM推理朝正确方向发展，有效充当误差校正器。本研究中，概率图模型通过包含所有验证任务的马尔可夫逻辑网络（MLN）实例化。

When using MLNs to model logical rules, various structures can be adopted depending on the task, including single-layer and double-layer configurations. For instance, in the case of Visual Relationship Detection (VRD), we employed a double-layer structure, as detailed in Section 5.4 and illustrated in Fig. 8. In other scenarios, we utilized a single-layer structure, with its joint probability distribution taking the form presented in Eq. (2). However, if the MLN incorporates multiple types of nodes and potential functions, the joint probability distribution will consist of multiple components. In this study, obtaining the nodes of the MLN requires performing grounding of the FOL statements. Grounding all FOLs in the database can lead to an excessively large number of variables, significantly increasing model complexity. Therefore, during training, the model identifies FOLs that are strongly related to the data, such as predicates that share the same labels as the data in a FOL. The optimization goal of the SRM is defined as O_{logic} in Eq. (4),which aims to maximize the joint probability distribution over all variables in terms of log-likelihood,
在使用MLN建模逻辑规则时，可根据任务采用不同结构，包括单层和双层配置。例如，在视觉关系检测（Visual Relationship Detection, VRD）中，我们采用了双层结构，详见第5.4节及图8。其他场景中采用单层结构，其联合概率分布形式如公式（2）所示。然而，若MLN包含多种节点类型和势函数，联合概率分布将由多个部分组成。本研究中，获取MLN节点需对一阶逻辑（First-Order Logic, FOL）语句进行归结。对数据库中所有FOL进行归结会导致变量数量过多，显著增加模型复杂度。因此，训练时模型识别与数据强相关的FOL，如与数据标签相同的谓词。SRM的优化目标定义为公式（4）中的O_{logic}，旨在最大化所有变量的联合概率分布的对数似然。

$$O_{\text{logic}} = \log P_{\theta_2, w}\left(\hat{y}, R\right) = \log\left\{\frac{1}{Z\left(w\right)}\exp\left\{\sum_{r\in R} w_r \sum_{a_r \in A_r} \phi\left(a_r\right) + \mathbb{C}\right\}\right\},$$

(4)

where $\mathbb{C}$ represents a custom term that may include potential functions $\phi_1, \phi_2, \ldots$,and should be designed according to task requirements.
其中$\mathbb{C}$表示自定义项，可能包含势函数$\phi_1, \phi_2, \ldots$，应根据任务需求设计。

8.4.5 4.2.3 Optimization

8.4.6 4.2.3 优化

The NSF-SRL model comprises two neural networks and a probabilistic graphical model, where the neural networks consist of a NRM and a concept network. The NRM is responsible for learning the features of concepts, while the concept network aims to infer the labels of query variables to approximate the posterior distribution. The symbolic reasoning module is responsible for learning a joint probability distribution to facillitate outcome inference.
NSF-SRL模型由两个神经网络和一个概率图模型组成，其中神经网络包括NRM（神经推理模块）和概念网络。NRM负责学习概念的特征，而概念网络旨在推断查询变量的标签以近似后验分布。符号推理模块负责学习联合概率分布以促进结果推断。

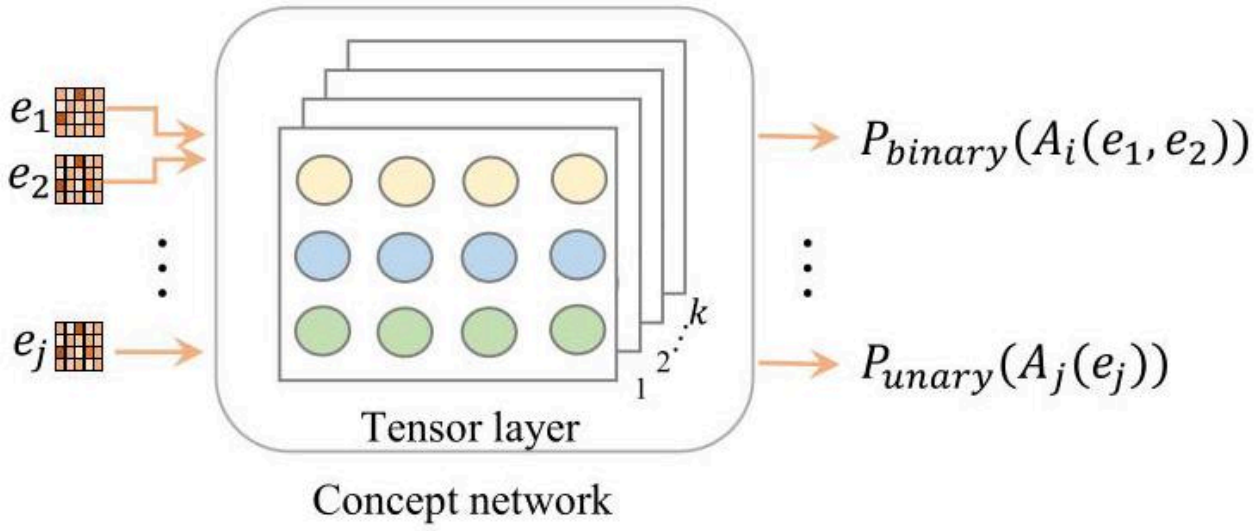


Fig. 3. Concept network. The inputs are feature vectors of object pairs (e.g., e_1 and e_2) or objects (e.g., e_j),and outputs are probabilities of affiliation relationship labels (e.g., $P_{\text{binary}}\left(A_i\left(e_1, e_2\right)\right)$) or object labels (e.g., $P_{\text{unary}}\left(A_j\left(e_j\right)\right)$). k represents tensor layer and each layer is a predicate.
图3. 概念网络。输入是对象对（例如e_1和e_2）或对象（例如e_j）的特征向量，输出是隶属关系标签的概率（例如$P_{\text{binary}}\left(A_i\left(e_1, e_2\right)\right)$）或对象标签（例如$P_{\text{unary}}\left(A_j\left(e_j\right)\right)$). k表示张量层，每层是一个谓词）。

The objective function $\log P_{\theta_1}$ of the neural reasoning module is typically differentiable and can be optimized using gradient descent. In this paper, the discrete logical knowledge within the symbolic reasoning module,represented as $\log P_{\theta_2, w}$,is transformed into a probabilistic graphical form, making symbolic reasoning also differentiable through the introduction of a concept network for posterior inference. The model aims to minimize the objective function to facilitate end-to-end joint training of both modules. Specifically, during the E-step, the posterior distribution of the

query variables is inferred, while in the M-step, the weights of the rules are learned. The training phase continues until the model reaches convergence. The parameters of the neural reasoning module, the concept network, and the symbolic reasoning module are denoted as θ_1, θ_2 ,and w ,respectively.
神经推理模块的目标函数$\log P_{\theta_1}$通常是可微的，可以通过梯度下降进行优化。本文中，符号推理模块内的离散逻辑知识表示为$\log P_{\theta_2,w}$，被转化为概率图形式，使符号推理通过引入概念网络进行后验推断也变得可微。模型旨在最小化目标函数以促进两个模块的端到端联合训练。具体地，在E步中推断查询变量的后验分布，在M步中学习规则的权重。训练阶段持续直到模型收敛。神经推理模块、概念网络和符号推理模块的参数分别表示为θ_1, θ_2和w。

To train the symbolic reasoning module, we need to maximize O_{logic} . However,the computation of the partition function $Z(w)$ in $P_{\theta_2,w}(\hat{y}, R)$ makes it intractable to optimize this objective function directly. Consequently, we introduce the variational EM algorithm and optimize the variational evidence lower bound (ELBO):
为了训练符号推理模块，我们需要最大化O_{logic}。然而，$P_{\theta_2,w}(\hat{y}, R)$中的配分函数$Z(w)$的计算使得直接优化该目标函数变得不可行。因此，我们引入变分EM算法并优化变分证据下界（ELBO）：

$$ELBO = E_Q\left[\log P_{\theta_2,w}(\hat{y}, R)\right] - E_Q\left[\log Q(\hat{y} \mid R)\right], \tag{5}$$

where $Q(\hat{y} \mid R)$ is the variational posterior distribution.
其中$Q(\hat{y} \mid R)$是变分后验分布。

In general, we utilize the variational EM algorithm to optimize the ELBO. Specifically, we minimize the Kullback-Leibler (KL) divergence between the variational posterior distribution $Q(\hat{y} \mid R)$ and the true posterior distribution $P_w(\hat{y} \mid R)$ during the E-step. Due to the complex graphical structure among variables, the exact inference becomes computationally intractable. Therefore, we adopt a mean-field distribution to approximate the true posterior, inferring the variables independently as follows:
通常，我们利用变分EM算法优化ELBO。具体来说，在E步中最小化变分后验分布$Q(\hat{y} \mid R)$与真实后验分布$P_w(\hat{y} \mid R)$之间的Kullback-Leibler（KL）散度。由于变量间复杂的图结构，精确推断计算上不可行。因此，我们采用均场分布近似真实后验，独立推断变量如下：

$$Q(\hat{y} \mid R) = \prod_{A_i \in A} Q(A_i). \tag{6}$$

For computational convenience, traditional variational methods typically require a predefined distribution, such as the Dirichlet distribution, and then utilize traditional search algorithms to solve the problem. In contrast, we employ neural networks (concept networks in this paper) to parameterize the variational calculation in Eq. (6). Consequently, the variational process transforms into a parameter learning process for the neural networks. As illustrated in Fig. 3, the neural network is called the concept network and is used to compute the posterior $Q(A_i)$. Thus, $Q(A_i)$ is rewritten as $Q_{\theta_2}(A_i)$.
为计算方便，传统变分方法通常需要预定义分布，如Dirichlet分布，然后利用传统搜索算法求解问题。相比之下，我们采用神经网络（本文中的概念网络）对公式(6)中的变分计算进行参数化。因此，变分过程转变为神经网络的参数学习过程。如图3所示，该神经网络称为概念网络，用于计算后验$Q(A_i)$。因此，$Q(A_i)$被重写为$Q_{\theta_2}(A_i)$。

Based on the above analysis, combined with Eq. 4 and Eq. 6), Eq. 5 is rewritten as:
基于上述分析，结合公式4和公式6，公式5被重写为：

$$ELBO = \sum_{r \in R} w_r \sum_{a_r \in A_r} E_{Q_{\theta_2}}\left[\phi(a_r)\right] - \log Z(w) \tag{7}$$

$$-E_{Q_{\theta_2}}\left[\sum_{A_i \in A} Q_{\theta_2}(A_i)\right] + \mathbb{C}.$$

In Fig. 3, to attain predicate labels of the hidden variables, we first feed feature vectors into concept network, such as feature vector of an object pair (e_1, e_2) or the feature vector of a single object e_j . Then,the concept network outputs a binary predicate label if provided with feature vectors of an object pair; otherwise, it outputs a unary predicate label. For example, when we input the feature vector of an image of a zebra into the concept network, it can output the predicate "zebra". Furthermore, to enhance the performance of the concept network through supervised information, we introduce a cross-entropy loss for optimization, which serves as a log-likelihood,
在图3中，为了获得隐藏变量的谓词标签，我们首先将特征向量输入概念网络，例如一对对象的特征向量(e_1, e_2)或单个对象的特征向量e_j。然后，概念网络在提供对象对的特征向量时输出二元谓词标签；否则，输出一元谓词标签。例如，当我们将斑马图像的特征向量输入概念网络时，它可以输出谓词“斑马(zebra)”。此外，为了通过监督信息提升概念网络的性能，我们引入了交叉熵损失进行优化，该损失作为对数似然函数，

$$O_{cro} = -\sum_{A_i \in A} Q_{\theta_2}(A_i) \log \hat{y}_i = -\log \Pi_{A_i \in A} \hat{y}_i^{\,Q_{\theta_2}(A_i)}. \tag{8}$$

Thus, the overall E-step objective function becomes:
因此，整体的E步目标函数变为：

$$O = \alpha O_{\text{task}} + \beta O_{\text{logic}} - \gamma O_{\text{cro}}, \tag{9}$$

where α, β and γ are the hyperparameter to control the weight. We maximize Eq. 9) to learn model parameters, the details are as follows:
其中α, β和γ是用于控制权重的超参数。我们通过最大化公式(9)来学习模型参数，具体如下：

$$\{\theta_1^*, \theta_2^*\} = \arg\max_{\theta_1, \theta_2} O. \tag{10}$$

In the M-step, the model learns the weights of the first-order logic rules. As we optimize these weights, the partition function $Z(w)$ in Eq. (4) is no longer constant,while $Q_{\theta 2}$ remains fixed. The partition function $Z(w)$ consists of an exponential number of terms, rendering direct optimization of the ELBO intractable. To solve this issue, we employ pseudo-log-likelihood [36] to approximate the ELBO, which is defined as follows:
在M步中，模型学习一阶逻辑规则的权重。随着我们优化这些权重，公式(4)中的配分函数$Z(w)$不再是常数，而$Q_{\theta 2}$保持不变。配分函数$Z(w)$包含指数级数量的项，导致直接优化ELBO不可行。为解决此问题，我们采用伪对数似然(pseudo-log-likelihood) [36]来近似ELBO，其定义如下：

$$P_w(\hat{y}, R) \simeq E_{Q_{\theta_2}}\left[\sum_{A_i \in A} \log P_w\left(A_i \mid MB_{A_i}\right)\right], \tag{11}$$

where MB_{A_i} represents Markov blanket of the ground atom A_i . For each rule r that connects A_i to its Markov blanket,we optimize weights w_r using gradient descent,and derivative is given by the following:
其中MB_{A_i}表示基元原子A_i的马尔可夫毯(Markov blanket)。对于每条将A_i与其马尔可夫毯连接的规则r，我们使用梯度下降优化权重w_r，其导数如下所示：

$$\bigtriangledown_{w_r} E_{Q_{\theta_2}}\left[\log P_w\left(A_i \mid MB_{A_i}\right)\right] \simeq \hat{y}_i - P_w\left(A_i \mid MB_{A_i}\right), \tag{12}$$

where $\hat{y}_i = 0$ or 1 if A_i is an observed variable,and $\hat{y}_i = Q_{\theta_2}(A_i)$ otherwise.
其中$\hat{y}_i = 0$为1，当A_i为观测变量时，否则为$\hat{y}_i = Q_{\theta_2}(A_i)$。

8.5 4.3 Concept Manipulation

8.6 4.3 概念操作

As mentioned in the overview of NSF-SRL, concept manipulation includes transductive and inductive concept manipulation methods. Consequently, we designed two corresponding approaches, as illustrated in Fig. 4 When the test data intersects with the training data, transductive concept manipulation employs the trained task network to predict results and utilizes probability graphical model to derive the FOLs corresponding to these predictions, providing explanations, as shown in Fig. 4 (a). In contrast, when the test data is disjoint from the training data, inductive concept manipulation uses the trained task network to extract data features. By introducing new FOLs to generalize the model for addressing new tasks, fuzzy logic reasoning is then applied to deduce the prediction results, as depicted in Fig. 4 (b).

如NSF-SRL概述中所述，概念操作包括传导式和归纳式概念操作方法。因此，我们设计了两种对应的方法，如图4所示。当测试数据与训练数据有交集时，传导式概念操作利用训练好的任务网络进行预测，并通过概率图模型推导与这些预测对应的一阶逻辑(FOLs)，以提供解释，如图4(a)所示。相反，当测试数据与训练数据不相交时，归纳式概念操作使用训练好的任务网络提取数据特征。通过引入新的FOLs来泛化模型以应对新任务，随后应用模糊逻辑推理推断预测结果，如图4(b)所示。

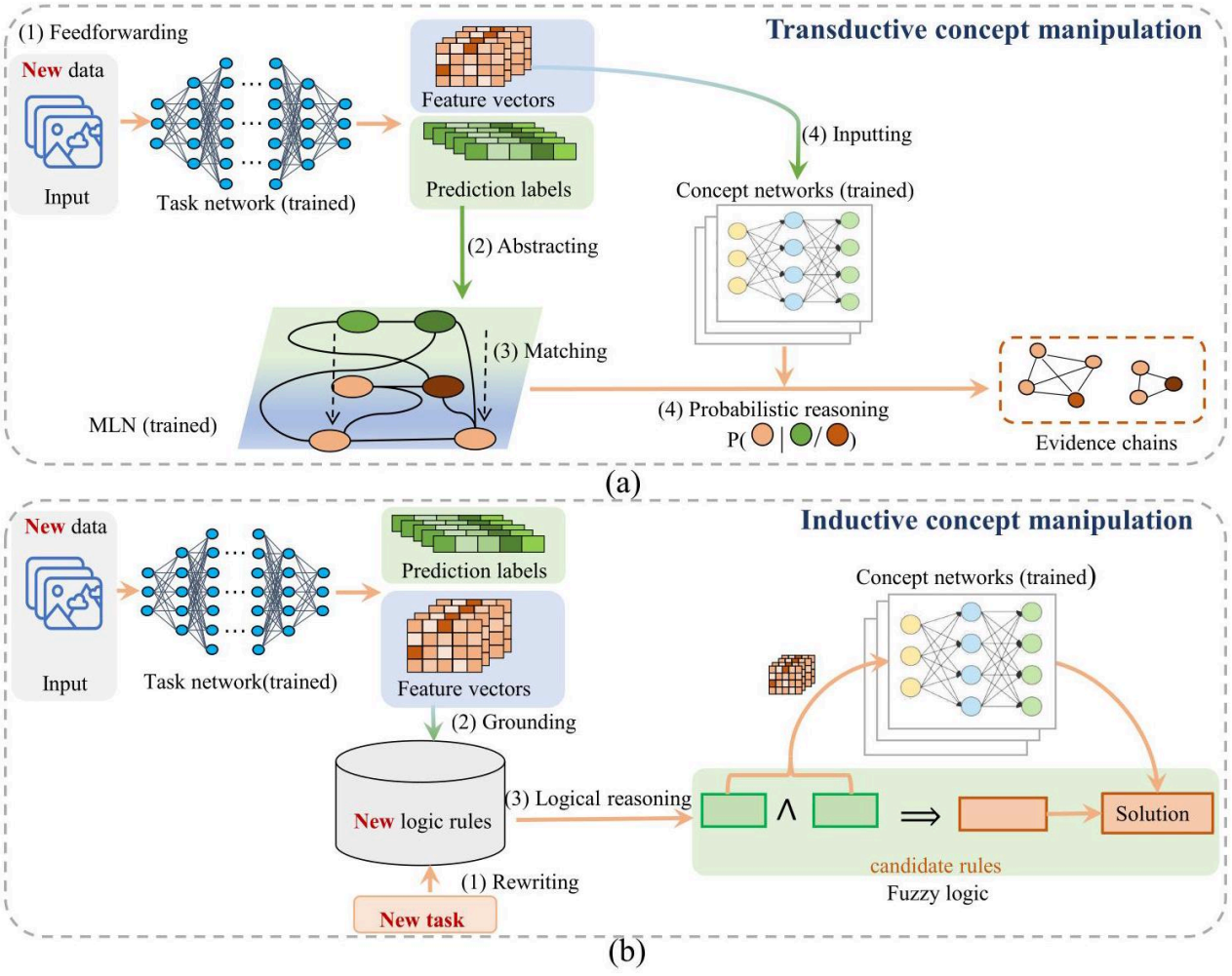


Fig. 4. Illustration of concept manipulation. (a) Transductive concept manipulation. The trained neural reasoning module predicts results, while the symbolic reasoning module provides interpretability. (b) Inductive concept manipulation. The trained neural reasoning module generates feature vectors, which are used by the symbolic reasoning module for reasoning.
图4. 概念操作示意图。(a) 传导式概念操作。训练好的神经推理模块进行预测，符号推理模块提供可解释性。(b) 归纳式概念操作。训练好的神经推理模块生成特征向量，符号推理模块基于此进行推理。

In the scenario depicted in Fig. 4 (a), the categories of the training set and the test set overlap. As illustrated in Fig. 1, the training data includes "zebra", which is also present in the test data. The steps of transductive manipulation are as follows: (1) Feedforwarding: input new data and obtain prediction labels and features through the trained task network; (2) Abstracting: derive partial nodes as observed variables in the probabilistic graphical model from the predicted labels; (3) Matching: match these partial nodes with first-order logic rules in the Markov logic network to identify candidate rules; (4) Inputting: feed feature vectors into the concept network, retrieve the scores of the concepts, and apply probabilistic reasoning (Eq. (13)) and fuzzy logic reasoning to obtain the probability score of each rule being true. Rules with high scores are selected as the evidence chain, interpreting the prediction labels. In this paper, we match the prediction results with ground atoms of the logic rules to achieve interpretability. A successful match indicates that the logic rules containing those ground atoms are triggered, and the corresponding clique composed of those nodes is selected. To quantify the likelihood that a candidate rule is true, we calculate the probability using t-norm fuzzy logic [37]. This process allows us to obtain evidence in the form of logic rules supporting the reasoning outcomes. To enhance interpretability, we select the most prominent piece of evidence in terms of a specific rule based on the posterior probability $P\left(r \mid \hat{y}\right)$ as follows:
在图4(a)所示的场景中，训练集和测试集的类别存在重叠。如图1所示，训练数据包含“斑马”(zebra)，测试数据中也存在该类别。传导操作的步骤如下：(1) 前馈：输入新数据，通过训练好的任务网络获得预测标签和特征；(2) 抽象：从预测标签中提取部分节点作为概率图模型中的观测变量；(3) 匹配：将这些部分节点与马尔可夫逻辑网络中的一阶逻辑规则匹配，以识别候选规则；(4) 输入：将特征向量输入概念网络，检索概念的得分，并应用概率推理（公式(13)）和模糊逻辑推理，得到每条规则为真的概率分数。得分较高的规则被选为证据链，用以解释预测标签。本文中，我们将预测结果与逻辑规则的基元原子(ground atoms)匹配以实现可解释性。匹配成功表明包含这些基元原子的逻辑规则被触发，且由这些节点组成的团(clique)被选中。为了量化候选规则为真的可能性，我们采用t-范数模糊逻辑[37]计算概率。该过程使我们能够获得以逻辑规则形式支持推理结果的证据。为增强可解释性，我们基于后验概率$P\left(r \mid \hat{y}\right)$选择特定规则中最显著的证据如下：

$$P\left(r \mid \hat{y}\right) = \prod_{A_i \in T_r} p\left(A_i \mid \hat{y}\right), \tag{13}$$

where T_r is the candidate rule here. Here, A_i is the ground atom sets in T_r .
其中T_r为此处的候选规则。这里，A_i是T_r中的基元原子集合。

In the scenario depicted in Fig. 4 (b), the categories of the training data and the test data do not overlap. As shown in Fig. 1 the training data does not include "tiger," whereas the test data does. Specifically, there are three steps for inductive manipulation: (1) Rewriting: rewriting logic rules based on the new task to accommodate specific requirements; (2) Grounding: grounding the logic rules using feature vectors from the task network; (3) Logic reasoning: inputting feature vectors of the concepts mentioned in the rule body of the candidate rules into the concept network to obtain the labels of the concepts. Subsequently, we reason the solution for the new task based on both the rule head and the rule body. This process can be seen as reprogramming for a new problem, utilizing the learned concepts from the previous step to tackle more complex problem scenarios. For instance, the model is trained on single-digit image addition and tested on multi-digit image addition tasks. By adopting this approach, the model can adapt its knowledge and reasoning capabilities to address new problems, thereby demonstrating the generalization capabilities of our method.

在图4(b)所示的场景中，训练数据和测试数据的类别不重叠。如图1所示，训练数据不包含“老虎”(tiger)，而测试数据中包含。具体而言，归纳操作包括三个步骤：(1) 重写：根据新任务重写逻辑规则以满足特定需求；(2) 归结：利用任务网络的特征向量对逻辑规则进行归结；(3) 逻辑推理：将候选规则规则体中提及的概念的特征向量输入概念网络，获得概念标签。随后，我们基于规则头和规则体推理新任务的解答。该过程可视为针对新问题的重新编程，利用前一步学习的概念来处理更复杂的问题场景。例如，模型在单数字图像加法任务上训练，在多数字图像加法任务上测试。通过此方法，模型能够调整其知识和推理能力以应对新问题，从而展示了我们方法的泛化能力。

9 5 EXPERIMENTS

10 5 实验

In this section, we conduct experiments on various tasks, including supervised task (transductive concept manipulation), weakly supervised task (transductive concept manipulation and inductive concept manipulation), and zero-shot learning task (inductive concept manipulation), using classic datasets for validation. We first describe the datasets and evaluation metrics. Then, we report the empirical results, including the performance, generalization, and interpretability across different tasks. Finally, we present ablation studies and hyperparameter analysis. The code is available at https://github.com/Dongranyu/NSF-SRL.
本节中，我们在多种任务上进行实验，包括监督任务（传导概念操作）、弱监督任务（传导概念操作和归纳概念操作）以及零样本学习任务（归纳概念操作），并使用经典数据集进行验证。首先介绍数据集和评估指标。然后报告经验结果，包括不同任务下的性能、泛化性和可解释性。最后，展示消融研究和超参数分析。代码可在https://github.com/Dongranyu/NSF-SRL获取。

10.1 5.1 Experimental Setup

10.2 5.1 实验设置

Tasks and datasets: For the supervised task, we validate our approach on visual relationship detection task. The corresponding datasets are Visual Relationship Detection(VRD) [38] and VG200 [39]. For the weakly supervised task, we conduct experiments on a digit image addition task, utilizing the handwritten digit dataset MNIST. For the zero-shot learning task, we employ image classification for validation, using the AwA2 [40] and CUB [41] datasets.
任务与数据集：对于监督任务，我们在视觉关系检测任务上验证方法。对应数据集为Visual Relationship Detection(VRD) [38]和VG200 [39]。对于弱监督任务，我们在数字图像加法任务上进行实验，使用手写数字数据集MNIST。对于零样本学习任务，我们采用图像分类进行验证，使用AwA2 [40]和CUB [41]数据集。

The VRD contains 5,000 images, with 4,000 images as training data and 1,000 images as testing data. There are 100 object classes and 70 predicates (relations). The VRD includes 37,993 relation annotations with 6,672 unique relations and 24.25 relationships per object category. This dataset contains 1,877 relationships in test set never occur in training set, thus allowing us to evaluate the generalization of our model in zero-shot prediction.
VRD包含5,000张图像，其中4,000张作为训练数据，1,000张作为测试数据。共有100个物体类别和70个谓词（关系）。VRD包含37,993条关系注释，6,672个唯一关系，每个物体类别平均24.25个关系。该数据集中测试集有1,877条关系在训练集中未出现，因此可用于评估模型在零样本预测中的泛化能力。

The VG200 contains 150 object categories and 50 predicates. Each image has a scene graph of around 11.5 objects and 6.2 relationships. 70% of the images is used for training and the remaining 30% is used for testing.
VG200包含150个物体类别和50个谓词。每张图像包含约11.5个物体和6.2个关系的场景图。70%的图像用于训练，剩余的30%用于测试。

The MNIST is a handwritten digit dataset and includes 0-9 digit images. In this paper, the task is to learn the "single-digit addition" formula given two MNIST images and a "addition" label. To implement the experiment on single-digit image addition, we randomly choose the initial feature of two digits to concat a tuple and take their addition as their labels. MNIST has 60,000 train sets and 10,000 test sets.

MNIST是一个手写数字数据集，包含0-9的数字图像。本文的任务是给定两张MNIST图像和一个“加法”标签，学习“单数字加法”公式。为实现单数字图像加法实验，我们随机选择两个数字的初始特征，将其拼接成元组，并将它们的和作为标签。MNIST包含60,000个训练集和10,000个测试集。

The AwA2 consists of 50 animal classes with 37,322 images. Training data contains 40 classes with 30,337 images, and test data has 10 classes with 6,985 images. Additionally, AwA2 provides 85 numeric attribute values for each class.
AwA2包含50个动物类别，共37,322张图像。训练数据包含40个类别，共30,337张图像，测试数据包含10个类别，共6,985张图像。此外，AwA2为每个类别提供85个数值属性值。

The CUB comprises 11,788 images spanning 200 bird classes, each associated with 312 attributes. Among these classes, 150 classes are designated as seen during training, while the remaining 50 are unseen and used for evaluation.
CUB包含11,788张图像，涵盖200个鸟类类别，每个类别关联312个属性。在这些类别中，150个类别在训练时可见，剩余50个类别为不可见，用于评估。

The logic rules. In this paper, logic rules encode relationships between a subject and multiple objects for visual relationship detection. Here, we build logic rules in an artificial way for VRD and VG200 datsets. That is, we take relationship annotations together with their subjects and objects to construct a logic rule according to the annotation file in the dataset. For example,we can obtain a logic rule as $\text{laptop}\,(x) \wedge$ next $\text{to}\,(x, y) \Rightarrow$ keyboard $(y) \vee$ mouse(y)by the above method. As a result,the numbers of logic rules are 1,642. Unlike VRD datasets, MNIST has no relationship annotation. To adapt to our weakly supervised task, we define corresponding logic rules, e.g., combining two single-digit labels and their addition label as logic rule. For example, $\text{digit}\,(x, d_1) \wedge \text{digit}\,(y, d_2) \Rightarrow$ addition $(d_1 + d_2, z)$,where the rule head is the addition label, and the rule body is two single-digit labels. In zero-shot image classification, we design logic rules for the AwA2 and CUB datasets, where the rule head is animal categories and the rule body consists of their attributes. For instance,catlike $(x) \wedge \text{tawny}\,(x) \wedge \text{spot}\,(x) \Rightarrow$ leopard(x).
逻辑规则。本文中，逻辑规则用于编码视觉关系检测中主体与多个客体之间的关系。这里，我们以人工方式为VRD和VG200数据集构建逻辑规则。即根据数据集中的标注文件，结合关系标注及其主体和客体构造逻辑规则。例如，通过上述方法可获得逻辑规则$\text{laptop}\,(x) \wedge$ next $\text{to}\,(x, y) \Rightarrow$ keyboard $(y) \vee$ mouse(y)。因此，逻辑规则数量为1,642条。与VRD数据集不同，MNIST无关系标注。为适应我们的弱监督任务，我们定义相应的逻辑规则，例如将两个单数字标签及其加法标签组合作为逻辑规则。例如，$\text{digit}\,(x, d_1) \wedge \text{digit}\,(y, d_2) \Rightarrow \text{addition}\,(d_1 + d_2, z)$，其中规则头为加法标签，规则体为两个单数字标签。在零样本图像分类中，我们为AwA2和CUB数据集设计逻辑规则，规则头为动物类别，规则体由其属性组成。例如，$\text{catlike}\,(x) \wedge \text{tawny}\,(x) \wedge \text{spot}\,(x) \Rightarrow$ leopard(x)。

Metrics: For VRD, we adopt evaluation metrics same as [42], which runs Relationship detection (ReD) and Phrase detection (PhD) and shows recall rates (Recall@) for the top 50/100 results, with $k = 1, 70$ candidate relations per relationship proposal (or k relationship predictions for per object box pair) before taking the top 50/100 results. ReD is inputting an image and outputting labels of triples and boxes of the objects. PhD is inputting an image and output labels and boxes of triples.
指标：对于VRD，我们采用与文献[42]相同的评估指标，执行关系检测（ReD）和短语检测（PhD），并展示前50/100结果的召回率（Recall@），在取前50/100结果前，每个关系提议有$k = 1, 70$个候选关系（或每对对象框有k个关系预测）。ReD输入图像，输出三元组标签及对象框。PhD输入图像，输出三元组的标签和框。

For VG200, we use the same evaluation metrics used in [42], including 1) Scene Graph Classification (SGCLS), which is to predict labels of the subject, object, and predicate given ground truth subject and object boxes; 2) Predicate Classification(PCLS), where predict predicate labels are given ground truth subject and object boxes and labels. Recall@ under the top $20/50/100$ predictions are reported.
对于VG200，我们使用文献[42]中相同的评估指标，包括1）场景图分类（SGCLS），即在给定真实主体和客体框的情况下预测主体、客体和谓词的标签；2）谓词分类（PCLS），在给定真实主体和客体框及标签的情况下预测谓词标签。报告前$20/50/100$预测的召回率。

For MNIST, AwA2 and CUB, we adopt accuracy(Acc) to evaluate the performance of the model. They are defined as Eq. (14).
对于MNIST、AwA2和CUB，我们采用准确率（Acc）来评估模型性能。其定义见公式（14）。

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{14}$$

where TP denotes true positive, TN denotes true negative, FP indicates false positive,and FN is false negative.
其中，TP表示真正例，TN表示真反例，FP表示假正例，FN表示假反例。

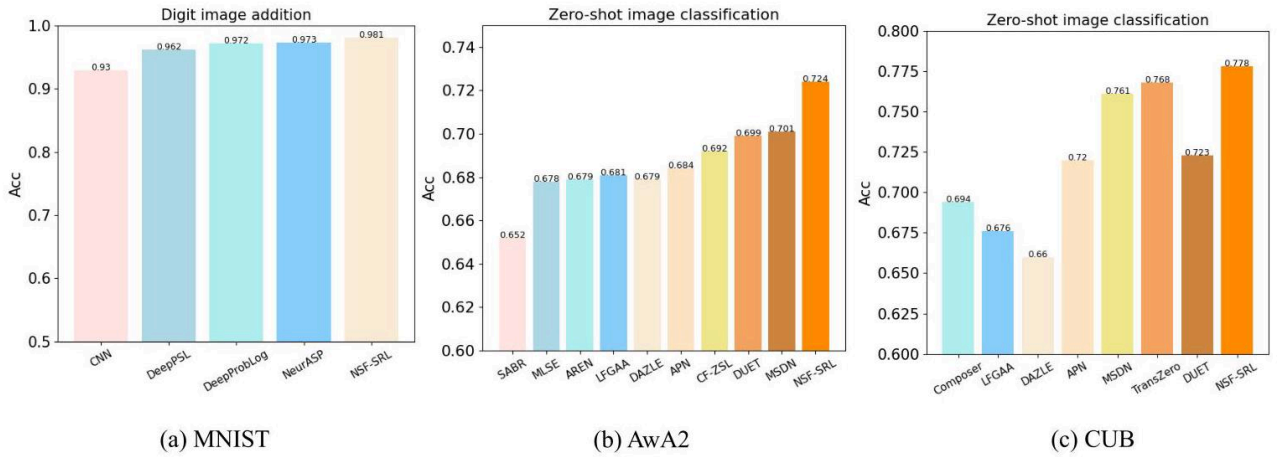
For the logic rule, we compute the probability of a logic rule that is true as an evaluation of logic rules. Here, we adopt Łukaseiwicz of t-norm fuzzy logic [37].
对于逻辑规则，我们计算逻辑规则为真的概率作为逻辑规则的评估。这里，我们采用t-范数模糊逻辑中的Łukasiewicz方法[37]。

10.3 5.2 Digit Image Addition Task

10.4 5.2 数字图像加法任务

In the context of neural-symbolic studies, digit image addition serves as a benchmark task, and MNIST dataset is recognized as a benchmark dataset. We evaluate the performance of our NSF-SRL model by comparing it against several neural-symbolic approaches and convolutional neural networks (CNNs). The neural-symbolic approaches considered include DeepPSL [43], DeepProbLog [12], and NeurASP [26]. This paper assesses the model's performance specifically on the single-digit image addition task, where two single-digit images are input into NSF-SRL, and the output is the predicted addition result. Furthermore, to verify the model's generalization capability, we also perform the multi-digit image addition task in Section 5.5.
在神经符号研究背景下，数字图像加法作为基准任务，MNIST数据集被公认为基准数据集。我们通过将NSF-SRL模型与多种神经符号方法及卷积神经网络（CNN）进行比较，评估其性能。所考虑的神经符号方法包括DeepPSL [43]、DeepProbLog [12]和NeurASP [26]。本文专门评估模型在单数字图像加法任务上的表现，该任务中将两个单数字图像输入NSF-SRL，输出为预测的加法结果。此外，为验证模型的泛化能力，我们还在第5.5节中进行了多数字图像加法任务的实验。

In the digit image addition task, the neural reasoning module first extracts image features using a CNN. These features are then processed through two fully connected layers to produce a 10-dimensional output vector. The activation functions employed in this neural network structure are ReLU and Softmax. We set the learning rate to 1e-4 and train the model for 15,000 epochs. Additionally, we utilize a batch size of 64 during training and a batch size of 1,000 during testing.
在数字图像加法任务中，神经推理模块首先使用CNN提取图像特征。然后，这些特征通过两个全连接层处理，生成一个10维的输出向量。该神经网络结构中采用的激活函数为ReLU和Softmax。我们将学习率设置为1e-4，训练模型15,000个周期。此外，训练时使用批量大小为64，测试时使用批量大小为1,000。



(a) MNIST (b) AwA2 (c) CUB

Fig. 5. Performance of NSF-SRL and comparison methods on digit image addition and zero-shot image classification tasks: (a) MNIST ; (b) AwA2 ; (c) CUB.
图5. NSF-SRL及对比方法在数字图像加法和零样本图像分类任务上的性能：(a) MNIST；(b) AwA2；(c) CUB。

Fig. 5 (a) presents the results of NSF-SRL alongside comparison methods for the digit image addition task. By comparing the performance of NSF-SRL with that of the other methods, we observe that NSF-SRL achieves decent performance. This finding underscores the feasibility of NSF-SRL in circumventing the reliance on strong supervised information typically required in conventional deep learning approaches. By integrating symbolic knowledge, NSF-SRL effectively leverages additional supervisory signals, such as data labels and relationships between data, resulting in improved model performance.
图5(a)展示了NSF-SRL及对比方法在数字图像加法任务中的结果。通过比较NSF-SRL与其他方法的性能，我们观察到NSF-SRL取得了良好的表现。该结果强调了NSF-SRL在规避传统深度学习方法中通常依赖强监督信息的可行性。通过整合符号知识，NSF-SRL有效利用了额外的监督信号，如数据标签和数据间关系，从而提升了模型性能。

10.5 5.3 Zero-shot Image Classification Task

10.6 5.3 零样本图像分类任务

In contrast to the digit image addition task, the zero-shot image classification task is inherently more complex. This task involves training a model on images of seen classes, enabling it to recognize images of unseen classes. The objective of the neural reasoning module in this context is to learn a mapping function from the visual space to the semantic space, thereby extracting image features of the objects. The symbolic reasoning module first receives these image features from the neural reasoning module, then models the logic rules using a MLN to learn the joint probability distribution. Finally, it employs a concept network to calculate the posterior probability of the joint distribution, predicting attribute labels and combining these labels according to the established logic rules.
与数字图像加法任务相比，零样本图像分类任务本质上更为复杂。该任务涉及在已见类别的图像上训练模型，使其能够识别未见类别的图像。神经推理模块的目标是在视觉空间到语义空间之间学习映射函数，从而提取对象的图像特征。符号推理模块首先接收来自神经推理模块的图像特征，然后利用马尔可夫逻辑网络（MLN）建模逻辑规则以学习联合概率分布。最后，符号推理模块通过概念网络计算联合分布的后验概率，预测属性标签，并根据既定逻辑规则组合这些标签。

In the zero-shot image classification task, the neural reasoning module is a CNN initialized with a pre-trained GoogleNet. Given an input image, we first use the CNN to extract initial visual features. These features are then fed into an attention network to attain discriminative image features. To enhance data augmentation, images undergo random cropping before being input into the model. For optimization, we employ the Adam optimizer with the following configurations: 15 epochs, a batch size of 64, and a learning rate of $1e-4$. The specific neural

architecture is illustrate in the Fig 6
在零样本图像分类任务中，神经推理模块为初始化了预训练GoogleNet的CNN。给定输入图像，首先使用CNN提取初始视觉特征。随后，这些特征被输入注意力网络以获得判别性图像特征。为增强数据增强效果，图像在输入模型前进行随机裁剪。优化方面，我们采用Adam优化器，配置为15个周期，批量大小64，学习率为1e − 4。具体神经架构如图6所示。

The symbolic reasoning module is implemented as a MLN, which integrates the neural reasoning module with FOL to extract discriminative image features. Additionally, this module enables the trained model to adapt from recognizing seen classes to unseen classes. Specifically, the symbolic reasoning module employs the MLN to learn the joint probability distribution between symbolic discriminative features and classes, predicting the labels of these features by calculating the posterior probability. Consequently, the symbolic reasoning module effectively combines the image features extracted by the neural reasoning module with FOL to perform fuzzy logic reasoning and derive class labels. The introduction of the MLN provides an efficient method for integrating visual features and symbolic discriminative features, thereby enhancing the model's generalization capability to unseen classes. This joint modeling approach captures the associations between image features and attributes, ultimately improving the model's performance in zero-shot image classification.
符号推理模块实现为马尔可夫逻辑网络（MLN），该模块将神经推理模块与一阶逻辑（FOL）结合以提取判别性图像特征。此外，该模块使训练好的模型能够从识别已见类别适应到未见类别。具体而言，符号推理模块利用MLN学习符号判别特征与类别之间的联合概率分布，通过计算后验概率预测这些特征的标签。因此，符号推理模块有效结合神经推理模块提取的图像特征与一阶逻辑，执行模糊逻辑推理并推导类别标签。MLN的引入为整合视觉特征与符号判别特征提供了高效方法，从而增强了模型对未见类别的泛化能力。该联合建模方法捕捉了图像特征与属性之间的关联，最终提升了模型在零样本图像分类中的表现。

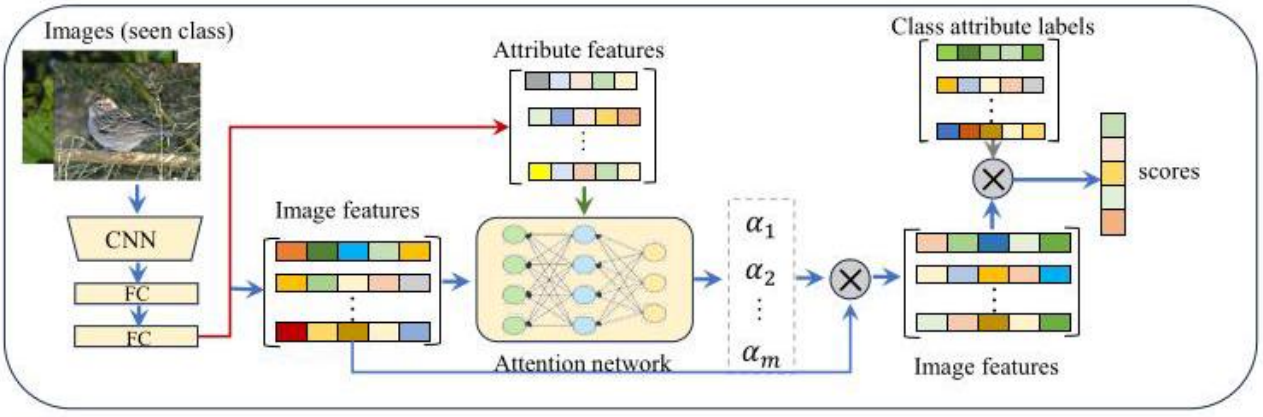


Fig. 6. The neural reasoning module on zero-shot image classification.
图6. 零样本图像分类中的神经推理模块。

Zero-shot image classification is a complex reasoning task that current neural-symbolic methods struggle to address effectively. Consequently, we primarily adopted deep learning-based contrastive approaches. Fig. 5 (b) presents the results for the AwA2 dataset, comparing our method against baseline methods such as SABR [44], MLSE [45], AREN [46], LFGAA [47], DAZLE [48], APN [49], CF-ZSL [50], DUET [51], and MSDN [52]. Fig. 5(c) presents the comparative results on the CUB dataset. The methods included in this comparison are Composer [53], LFGAA [47], DAZLE [48], APN [49], MSDN [52], TransZero [54], and DUET [51].
零样本图像分类是一项复杂的推理任务，当前神经符号方法难以有效解决。因此，我们主要采用基于深度学习的对比

方法。图5(b)展示了AwA2数据集上的结果，将我们的方法与基线方法如SABR [44]、MLSE [45]、AREN [46]、LFGAA [47]、DAZLE [48]、APN [49]、CF-ZSL [50]、DUET [51]和MSDN [52]进行了比较。图5(c)展示了CUB数据集上的比较结果，所比较的方法包括Composer [53]、LFGAA [47]、DAZLE [48]、APN [49]、MSDN [52]、TransZero [54]和DUET [51]。

From Fig. 5, it is evident that our NSF-SRL achieves optimal performance across different datasets, validating the effectiveness of the model. This success can be attributed to the logical rules that model the relationships between attribute features, seen categories, and unseen categories, including co-occurrence relationships. Such rules facilitate the model in capturing these relationships, thereby enhancing classification performance. Additionally, this experiment highlights that incorporating symbolic reasoning with FOL enhances the robustness of the model.
从图5可以看出，我们的NSF-SRL在不同数据集上均取得了最佳性能，验证了该模型的有效性。这一成功归因于建模属性特征、已见类别和未见类别之间关系的逻辑规则，包括共现关系。这些规则帮助模型捕捉这些关系，从而提升分类性能。此外，该实验还表明，结合一阶逻辑(FOL)的符号推理增强了模型的鲁棒性。

10.7 5.4 Visual Relationship Detection Task

10.8 5.4 视觉关系检测任务

Visual relationship detection, similar to zero-shot image classification, is a complex task that aims to identify objects within an image and the relationships between them. These relationships can be represented as triplets (subject, predicate, object). In this context, the neural reasoning module serves as a deep learning-based specifically designed for visual relationship detection, extracting label concepts of both objects and their relationships from the input image. Conversely, the symbolic reasoning module functions as a two-layer probabilistic graphical model, intended to integrate the learned object and relationship labels while guiding the learning process of the visual reasoning module.
视觉关系检测类似于零样本图像分类，是一项复杂任务，旨在识别图像中的对象及其之间的关系。这些关系可表示为三元组（主语，谓语，宾语）。在此背景下，神经推理模块作为一种基于深度学习的专门设计，用于视觉关系检测，从输入图像中提取对象及其关系的标签概念。相反，符号推理模块作为一个两层概率图模型，旨在整合学习到的对象和关系标签，同时指导视觉推理模块的学习过程。

For the visual relationship detection task, our neural reasoning module is based on the architecture described in [42]. It consists of two components: a visual module and a semantic module. The visual module primarily extracts visual features using a CNN, specifically employing layers conv1_1 to conv5_3 of VGG16 to generate a global feature map of the image. Subsequently, the subject, relation, and object features are region-of-interest (ROI) pooled and processed through two fully connected layers to produce three intermediate hidden features. The semantic module, on the other hand, processes word vectors corresponding to the subject, relation, and object labels via a multilayer perceptron (MLP) to generate embeddings. Before training, we initialize each branch using pre-trained weights from the COCO dataset [55] and adopt word2vec [56] for the word vectors in our experiments. Specifically, we train our model for 7 epochs with a learning rate set to 1e-4, and the dimension of the object feature is established at 512. The specific neural architecture is illustrated in Fig. 7
对于视觉关系检测任务，我们的神经推理模块基于文献[42]中描述的架构。它由两个部分组成：视觉模块和语义模块。视觉模块主要利用卷积神经网络(CNN)提取视觉特征，具体采用VGG16的conv1_1至conv5_3层生成图像的全局特征图。随后，对主语、关系和宾语特征进行感兴趣区域(ROI)池化，并通过两个全连接层处理，生成三个中间隐藏特征。语义模块则通过多层感知机(MLP)处理对应主语、关系和宾语标签的词向量，生成嵌入。在训练前，我们使用COCO数据集[55]的预训练权重初始化各分支，并在实验中采用word2vec[56]作为词向量。具体而言，我们以学习率1e-4训练模型7个epoch，对象特征维度设为512。具体神经架构见图7。

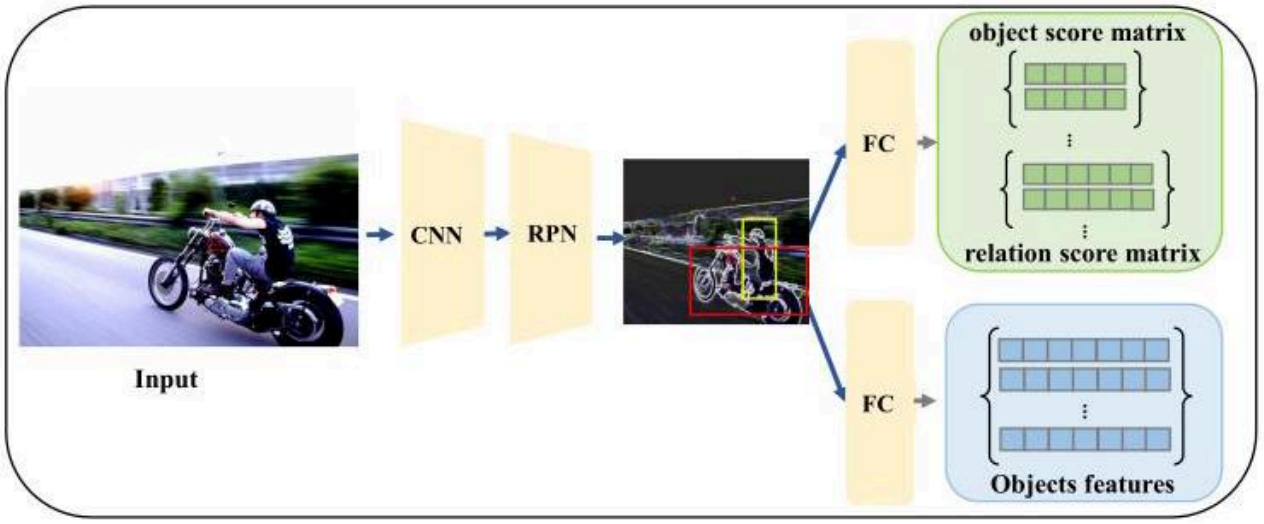


Fig. 7. The neural reasoning module on visual relationship detection task.
图7. 视觉关系检测任务中的神经推理模块。

As illustrated in the Fig. 8, the symbolic reasoning module is structured as a bi-level probabilistic graphical model, where the high-level layer represents the prediction results (pseudo-labels) generated by the neural reasoning module. In contrast, the low-level layer consists of the ground atoms of MLN. This module consists of two types of nodes (random variables) and cliques (potential functions): the prediction labels from the neural reasoning module in the high-level layer nodes and the ground atoms of the MLN in the low-level layer nodes. Let $\hat{y} = \{\widehat{y_1}, \widehat{y_2}, \ldots\}$ denote the set of high-level nodes (pseudo-labels),and let $A = \{A_1, A_2, \ldots\}$ represent the set of low-level nodes, comprising the ground atoms in the FOLs. A clique $\{\hat{y}_i, A_j\}$ signifies the correlation between these levels,while another clique A_r represents the ground atoms of a FOL. Consequently,the custom term $\mathbb{C}$ can be defined as $\sum_{\widehat{y}_i \in \hat{y}, A_j \in A} \phi_1\left(\widehat{y}_i, A_j\right)$ in Eq. 4).

如图8所示，符号推理模块构建为双层概率图模型，高层表示神经推理模块生成的预测结果（伪标签），低层由马尔可夫逻辑网络(MLN)的基元原子组成。该模块包含两类节点（随机变量）和团（势函数）：高层节点为神经推理模块的预测标签，低层节点为MLN的基元原子。设$\hat{y} = \{\widehat{y_1}, \widehat{y_2}, \ldots\}$表示高层节点集合（伪标签），$A = \{A_1, A_2, \ldots\}$表示低层节点集合，包括一阶逻辑(FOL)中的基元原子。团$\{\hat{y}_i, A_j\}$表示这两层之间的相关性，另一团A_r表示FOL的基元原子。因此，自定义项$\mathbb{C}$可定义为公式(4)中的$\sum_{\widehat{y}_i \in \hat{y}, A_j \in A} \phi_1\left(\widehat{y}_i, A_j\right)$。

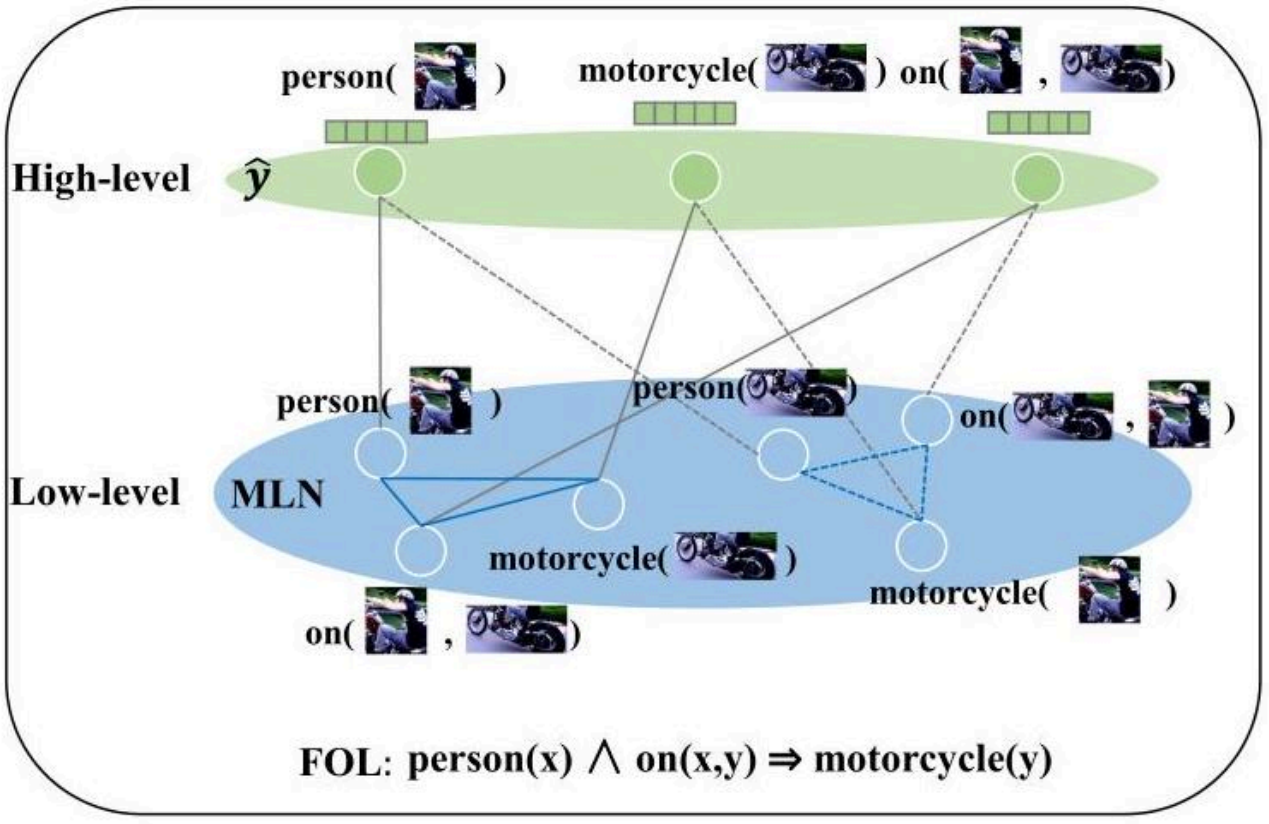


Fig. 8. The symbolic reasoning module on visual relationship detection task.
图8. 视觉关系检测任务中的符号推理模块。

Existing neural-symbolic methods, such as DeepProbLog, have not been validated on complex tasks like visual relationship detection. Therefore, our comparative methods are restricted to those based solely on deep learning. The experimental results of NSF-SRL and several comparative methods are presented in Table 2 for VRD dataset. As not all comparative methods specified k in their experiment,we report results as "free k " when treating k as a hyperparameter. The results indicate that our NSF-SRL outperforms the comparative methods in most cases. The enhancements offered by the symbolic reasoning module can be attributed to two key factors. First, the symbolic reasoning module is structured as a probabilistic graphical model that effectively captures dependencies between variables, facilitating a more accurate modeling of complex relationships. Second, our logic rules are constructed based on the co-occurrence relationships among predicates, suggesting that when one object is present, another is likely to appear as well. By maximizing the joint probability of the probabilistic graphical model, we effectively enhance the co-occurrence probability during the training phase.
现有的神经符号方法，如DeepProbLog，尚未在视觉关系检测等复杂任务上得到验证。因此，我们的对比方法仅限于基于深度学习的方法。NSF-SRL及若干对比方法在VRD数据集上的实验结果见表2。由于并非所有对比方法在实验中都指定了k，我们在将k视为超参数时，将结果标注为“free k”。结果表明，NSF-SRL在大多数情况下优于对比方法。符号推理模块带来的提升可归因于两个关键因素。首先，符号推理模块被构建为概率图模型，有效捕捉变量间的依赖关系，从而更准确地建模复杂关系。其次，我们的逻辑规则基于谓词间的共现关系构建，表明当一个对象出现时，另一个对象也很可能出现。通过最大化概率图模型的联合概率，我们在训练阶段有效提升了共现概率。

Table 3 presents the results on the VG200 dataset. Notably, the state-of-the-art methods do not specify a clear value for k in this context. Therefore, we report the performance of our NSF-SRL model with $k = 1$. Our results demonstrate that NSF-SRL outperforms existing methods across two metrics in Recall@20/50/100, highlighting the advantages of leveraging symbolic knowledge through logic rules. Furthermore, while PCLS emphasizes relationship

recognition, NSF-SRL achieves a superior score on the PCLS evaluation metric, indicating that the incorporation of logic rules enhances relationship recognition capabilities within the model.
表3展示了VG200数据集上的结果。值得注意的是，当前最先进的方法在此情境下未明确指定k的具体值。因此，我们报告了NSF-SRL模型在$k = 1$条件下的性能。结果显示，NSF-SRL在Recall@20/50/100两个指标上均优于现有方法，凸显了通过逻辑规则利用符号知识的优势。此外，尽管PCLS强调关系识别，NSF-SRL在PCLS评估指标上取得了更高分数，表明逻辑规则的引入提升了模型的关系识别能力。

10.9 5.5 Generalization

10.10 5.5 泛化能力

Evaluating a model's generalization ability is essential, as it reflects its adaptability and robustness across diverse scenarios. In this study, generalization refers to the model's predictive performance on unseen samples. For example, the model is initially trained on a single-digit image addition task and subsequently tested on a multi-digit image addition task. Zero-shot image classification serves as an experiment that validates the model's generalization capabilities. Consequently, we only focus our experimental validation on visual relationship detection and digit image addition tasks.
评估模型的泛化能力至关重要，因为它反映了模型在不同场景下的适应性和鲁棒性。本研究中，泛化指模型对未见样本的预测性能。例如，模型最初在单数字图像加法任务上训练，随后在多数字图像加法任务上测试。零样本图像分类作为验证模型泛化能力的实验。因此，我们的实验验证仅聚焦于视觉关系检测和数字图像加法任务。

TABLE 2

Test performance of visual relationship detection. The recall results for the top 50/100 in "ReD" and "PhD" are reported, respectively. The best result is highlighted in bold. "-" denotes the corresponding result is not provided.
视觉关系检测的测试性能。分别报告了“ReD”和“PhD”中前50/100的召回率结果。最佳结果以粗体标出。“-”表示未提供对应结果。

Methods	ReD		PhD		ReD				PhD			
	free k				$k = 1$		$k = \{70\}$		$k = 1$		$k = \{70\}$	
Recall@	50	100	50	100	50	100	50	100	50	100	50	100
Lk distilation [57]	22.7	31.9	26.5	29.8	19.2	21.3	22.7	31.9	23.1	24.0	26.3	29.4
Zoom-Net [58]	21.4	27.3	29.1	37.3	18.9	21.4	21.4	27.3	28.8	28.1	29.1	37.3
CAI+SCA-M [58]	22.3	28.5	29.6	38.4	19.5	22.4	22.3	28.5	25.2	28.9	29.6	38.4
MF-URLN [59]	23.9	26.8	31.5	36.1	23.9	26.8	-	-	23.9	26.8	-	-
LS-VRU [42]	27.0	32.6	32.9	39.6	23.7	26.7	27.0	32.6	28.9	32.9	32.9	39.6
GPS-Net [60]	27.8	31.7	33.8	39.2	-	-	27.8	31.7	-	-	33.8	39.2
UVTransE [61]	27.4	34.6	31.8	40.4	25.7	29.7	27.3	34.1	30.0	36.2	31.5	39.8
NMP [62]	21.5	27.5	-	-	20.2	24.0	21.5	27.5	-	-	-	-
NSF-SRL	29.4	35.3	36.2	43.0	26.2	29.4	29.4	35.3	32.3	36.4	36.2	43.0

方法	ReD		博士（PhD）		ReD				博士（PhD）			
	免费 k				$k = 1$		$k = {70}$		$k = 1$		$k = {70}$	
召回率@	50	100	50	100	50	100	50	100	50	100	50	100
Lk蒸馏 [57]	22.7	31.9	26.5	29.8	19.2	21.3	22.7	31.9	23.1	24.0	26.3	29.4
Zoom-Net [58]	21.4	27.3	29.1	37.3	18.9	21.4	21.4	27.3	28.8	28.1	29.1	37.3
CAI+SCA-M [58]	22.3	28.5	29.6	38.4	19.5	22.4	22.3	28.5	25.2	28.9	29.6	38.4
MF-URLN [59]	23.9	26.8	31.5	36.1	23.9	26.8	-	-	23.9	26.8	-	-
LS-VRU [42]	27.0	32.6	32.9	39.6	23.7	26.7	27.0	32.6	28.9	32.9	32.9	39.6
GPS-Net [60]	27.8	31.7	33.8	39.2	-	-	27.8	31.7	-	-	33.8	39.2
UVTransE [61]	27.4	34.6	31.8	40.4	25.7	29.7	27.3	34.1	30.0	36.2	31.5	39.8
NMP [62]	21.5	27.5	-	-	20.2	24.0	21.5	27.5	-	-	-	-
NSF-SRL	29.4	35.3	36.2	43.0	26.2	29.4	29.4	35.3	32.3	36.4	36.2	43.0

10.10.1 5.5.1 Visual Relationship Detection

10.10.2 5.5.1 视觉关系检测

We evaluated the performance of our NSF-SRL model against the baseline LS-VRU in a zero-shot learning scenario. In this context, the training and testing data comprise disjoint sets of relationships from the VRD dataset, as illustrated in Fig. 9 (a). The results demonstrate that NSF-SRL outperforms LS-VRU across various recall metrics, highlighting LS-VRU's limitations in handling sparse relationships. In contrast, NSF-SRL effectively incorporates symbolic knowledge and language priors, making it less susceptible to the challenges posed by sparse relationships.
我们在零样本学习场景下评估了NSF-SRL模型相较于基线LS-VRU的性能。在此情境中，训练和测试数据由VRD数据集中不相交的关系集合组成，如图9(a)所示。结果表明，NSF-SRL在多种召回指标上均优于LS-VRU，凸显了LS-VRU在处理稀疏关系时的局限性。相比之下，NSF-SRL有效地融合了符号知识和语言先验，使其不易受到稀疏关系带来的挑战影响。

10.10.3 5.5.2 Digit image Addition

10.10.4 5.5.2 数字图像加法

We validate the generalization capability of NSF-SRL in multi-digit task by comparing it to the baseline. In multi-digit image addition, the input consists of two lists of images, each representing a digit, with each list corresponding to a multi-digit number. The label reflects the sum of these two numbers. In our experiment, a CNN is trained on the multi-digit image addition dataset to test the multi-digit image addition task, while we apply the learned model from the single-digit image addition task to this scenario. As shown in Fig. 9 (b), the results illustrate the enhanced prediction accuracy in the multi-digit image addition task by leveraging concepts acquired during the single-digit task.Our findings indicate a significant improvement compared to other methods, underscoring the flexibility of our model, which can generalize from simpler tasks to more complex ones by adapting its logic rules. Notably, this generalization is facilitated by the shared learnable concepts between the two tasks.
我们通过与基线模型的比较验证了NSF-SRL在多位数任务中的泛化能力。在多位数图像加法中，输入由两组图像列表组成，每个图像代表一个数字，每组列表对应一个多位数。标签反映这两个数字的和。在我们的实验中，训练了一个卷积神经网络（CNN）用于多位数图像加法数据集，以测试多位数图像加法任务，同时我们将单位数图像加法任务中学得的模型应用于该场景。如图9(b)所示，结果展示了通过利用单数任务中获得的概念，提升了多位数图像加法任务的预测准确率。我们的发现表明，与其他方法相比有显著提升，强调了我们模型的灵活性，能够通过调整其逻辑规则从简单任务泛化到更复杂任务。值得注意的是，这种泛化得益于两任务间共享的可学习概念。

10.11 5.6 Interpretibility

10.12 5.6 可解释性

We employ visual relationship detection and zero-shot image classification tasks to demonstrate the interpretability of our results. In the context of visual relationship detection, Fig. 10 (a) illustrates the reasoning behind the identified relationship "next to" between "laptop" and either the "keyboard" or "mouse". According to Eq. (13), when the subject is a "laptop" and the object is either "keyboard" or "mouse", the relationship "next to" is assigned the highest confidence by the logic rule laptop $(x) \wedge$ next to $(x, y) \Rightarrow$ keyboard $(y) \vee$ mouse(y).
我们采用视觉关系检测和零样本图像分类任务来展示结果的可解释性。在视觉关系检测的背景下，图10(a)展示了“笔记本电脑”与“键盘”或“鼠标”之间识别出“旁边”关系的推理过程。根据公式(13)，当主体为“笔记本电脑”，客体为“键盘”或“鼠标”时，逻辑规则 laptop $(x) \wedge$ next to $(x, y) \Rightarrow$ keyboard $(y) \vee$ mouse(y) 赋予“旁边”关系最高置信度。

TABLE 3

Comparative results for top 50/100 in "SGCLS" and "PCLS" respectively on the VG200 dataset. The best result is highlighted in bold.
在VG200数据集上，“SGCLS”和“PCLS”任务的前50/100名比较结果。最佳结果以粗体显示。

Metrics Recall@	SGCLS			PCLS		
Methods	20	50	100	20	50	100
VRD [38]	-	11.8	14.1	-	27.9	35.0
Ass-Embedding [63	18.2	21.8	22.6	47.9	54.1	55.4
Mess-Passing [39	31.7	34.6	35.4	52.7	59.3	61.3
Graph-RCNN [64	-	29.6	31.6	-	54.2	59.1
Per-Invariant [65	-	36.5	38.8	-	65.1	66.9
Motifnet | 66	32.9	35.8	36.5	58.5	65.2	67.1
LS-VRU [42]	36.0	36.7	36.7	66.8	68.4	68.4
GPS-Net 60	36.1	39.2	40.1	60.7	66.9	68.8
$\mathrm{{NSF}} - \mathrm{{SRL}}\left({k = 1}\right)$	37.0	39.3	39.3	67.8	69.1	70.0
指标 召回率@	SGCLS			PCLS		
方法	20	50	100	20	50	100
VRD [38]	-	11.8	14.1	-	27.9	35.0
Ass-Embedding [63	18.2	21.8	22.6	47.9	54.1	55.4
Mess-Passing [39	31.7	34.6	35.4	52.7	59.3	61.3
Graph-RCNN [64	-	29.6	31.6	-	54.2	59.1
Per-Invariant [65	-	36.5	38.8	-	65.1	66.9
Motifnet | 66	32.9	35.8	36.5	58.5	65.2	67.1
LS-VRU [42]	36.0	36.7	36.7	66.8	68.4	68.4
GPS-Net 60	36.1	39.2	40.1	60.7	66.9	68.8
$\mathrm{{NSF}} - \mathrm{{SRL}}\left({k = 1}\right)$	37.0	39.3	39.3	67.8	69.1	70.0

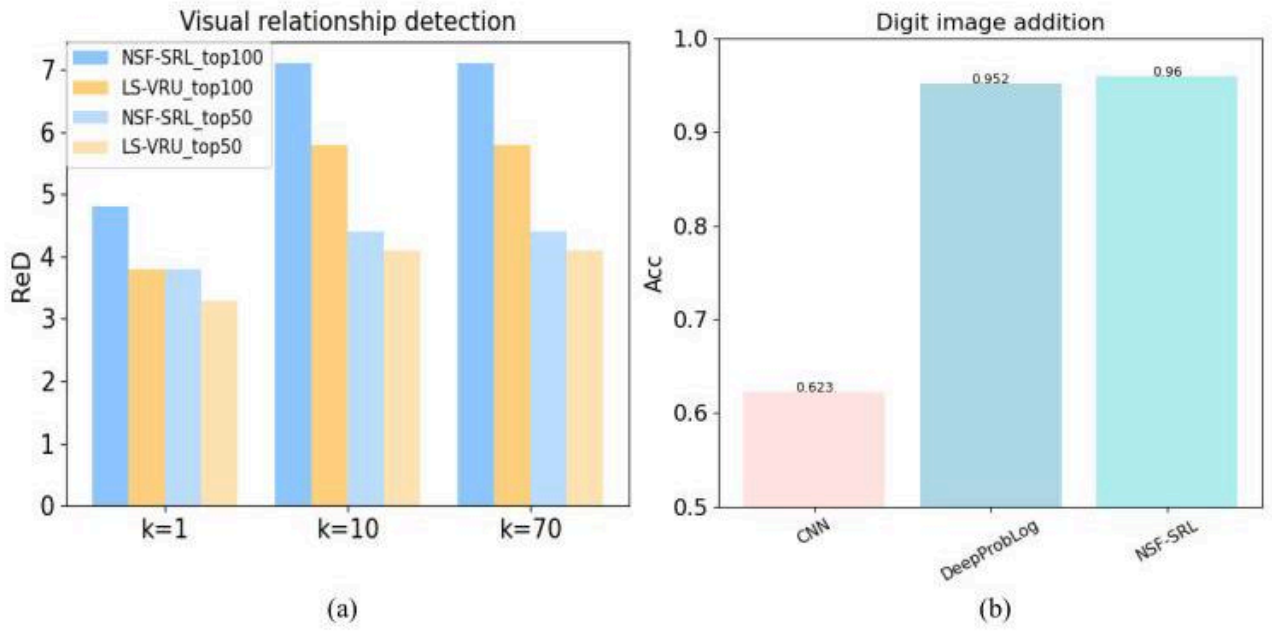


Fig. 9. Generalization of NSF-SRL and comparison methods on visual relationship detection and digit image addition tasks. (a) Visual relationship detection. Larger ReD indicates better results. (b) Multi-digit image addition.

图9. NSF-SRL及比较方法在视觉关系检测和数字图像加法任务上的泛化能力。(a) 视觉关系检测。ReD值越大表示结果越好。(b) 多位数字图像加法。

In zero-shot image classification, we used heatmaps to visualize the discriminative image features. As shown in Fig. 10 (b), the highlighted regions represent the discriminative features captured by our model. By combining the predicted discriminative feature labels with the logic rules, the model can infer class labels. This transparent reasoning process facilitates easy understanding of the model's decision-making when presented with an image. For instance, when the model identifies an image as black_billed_vuckoo, it justifies its prediction by highlighting features such as a curved_bill, tapered_wing and pointed_tail in the image, and logically deduces that the object possessing these features belongs to the black_billed_vuckoo class, based on the applied rule.

在零样本图像分类中，我们使用热力图可视化判别性图像特征。如图10(b)所示，突出显示的区域代表模型捕捉到的判别性特征。通过将预测的判别性特征标签与逻辑规则结合，模型能够推断类别标签。这一透明的推理过程便于理解模型在给定图像时的决策过程。例如，当模型将图像识别为黑嘴杜鹃(black_billed_cuckoo)时，它通过突出显示图像中的弯曲的喙(curved_bill)、锥形的翅膀(tapered_wing)和尖尾(pointed_tail)等特征，并基于应用的规则逻辑推断出具有这些特征的对象属于黑嘴杜鹃类，从而为其预测提供了合理解释。

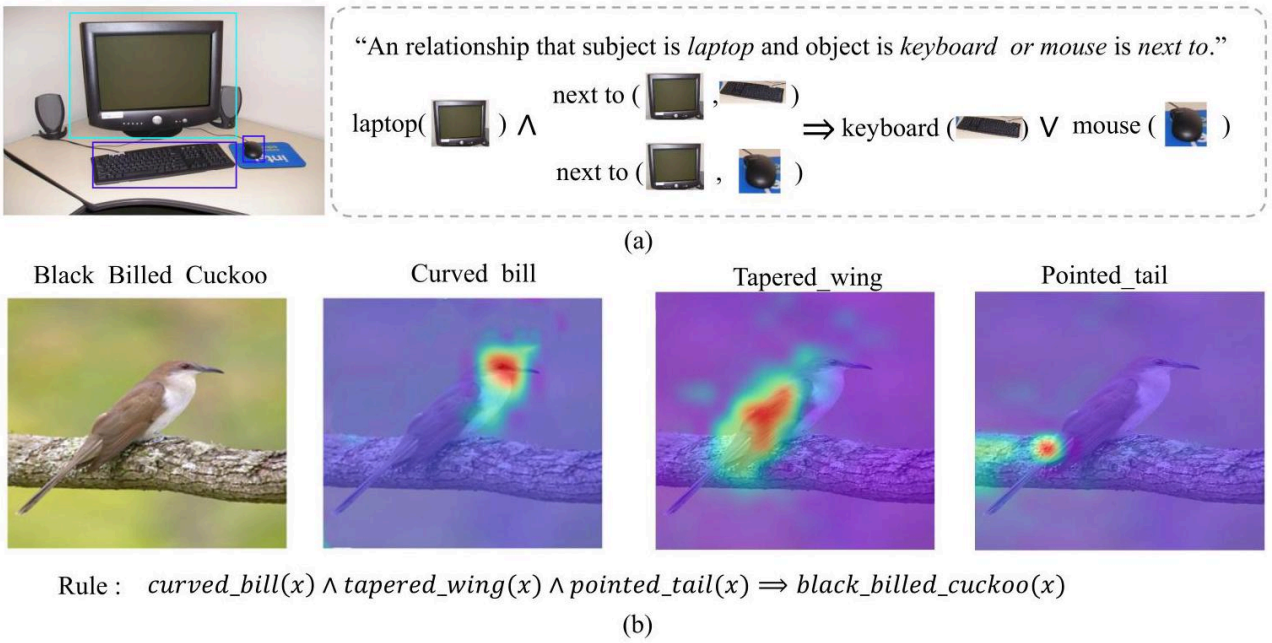


Fig. 10. Interpretability analysis. (a) An example illustrating the interpretability of NSF-SRL. For example, why is the relationship "next to" detected between a "laptop" and a "keyboard" or "mouse" in an image? According to Eq. (13), the model identifies the most confident logic rule: laptop(x) ∧ next to (x, y) ⇒ keyboard (y)∨ mouse(y). This demonstrates that the reasoning results of NSF-SRL align with common sense. (b) Visualization of the learned discriminative image features by our model. Key features, such as the shape of the bill, wing, and tail, are highlighted, providing a visual explanation of the model's reasoning.

图10. 可解释性分析。(a) 以NSF-SRL的可解释性为例说明。例如，为什么在图像中检测到"笔记本电脑(laptop)"与"键盘(keyboard)"或"鼠标(mouse)"之间存在"相邻(next to)"关系？根据公式(13)，模型识别出最有信心的逻辑规则：laptop(x) ∧ next to (x, y) ⇒ keyboard (y)∨ mouse(y)。这表明NSF-SRL的推理结果符合常识。(b) 我们模型学习到的判别性图像特征的可视化。关键特征如喙、翅膀和尾巴的形状被突出显示，为模型的推理提供了视觉解释。

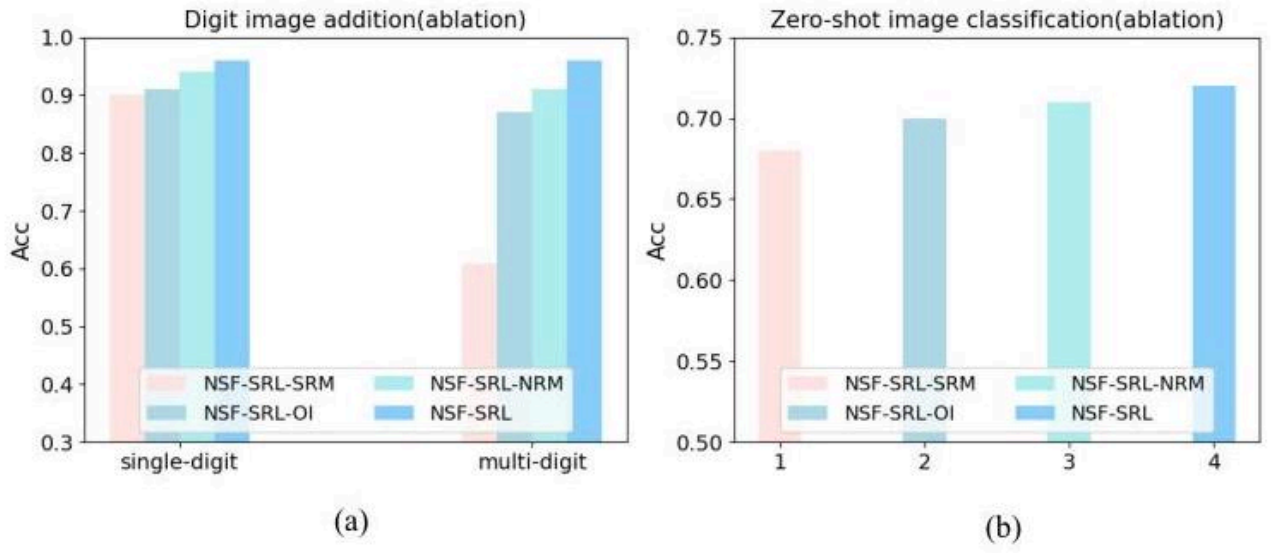


Fig. 11. Ablation results on digit image addition and zero-shot image classification tasks.

图11. 数字图像加法和零样本图像分类任务的消融实验结果。

10.13 5.7 Ablation Studies

10.14 5.7 消融研究

During the training phase, we conduct an extensive analysis of various factors that may affect downstream task performance. These factors include the hyperparameters α, β, γ . This comprehensive evaluation framework provides deeper insights into the influence of these factors on model performance.
在训练阶段，我们对可能影响下游任务性能的各种因素进行了广泛分析。这些因素包括超参数α, β, γ。这一全面的评估框架为深入理解这些因素对模型性能的影响提供了依据。

To investigate the impact of model trade-offs on reasoning, we designed three variants to assess the effect of individual components on NSF-SRL. Specifically, we derived these variants from the optimized objective in Eq. (9) by adjusting the values of the trade-off factors. The three variants are as follows: (1) NSF-SRL-SRM $(\alpha = 1, \beta = 0, \gamma = 0)$: excluding the symbolic reasoning module,(2) NSF-SRL-NRM $(\alpha = 1/2, \beta = 1, \gamma = 1)$: reducing the visual reasoning module by half, and (3) NSF-SRL-OI $(\alpha = 1, \beta = 1, \gamma = 0)$: omitting the cross-entropy of observed variables. We conducted experiments on digit image addition and zero-shot image classification tasks to evaluate performance of NSF-SRL and its variants. The results are presented in Fig. 11 (a) and Fig. 11(b), respectively.
为探究模型权衡对推理的影响，我们设计了三个变体以评估各组成部分对NSF-SRL的作用。具体而言，我们通过调整权衡因子的取值，从公式(9)的优化目标中派生出这三个变体。三种变体分别为：(1) NSF-SRL-SRM $(\alpha = 1, \beta = 0, \gamma = 0)$ ： 排除符号推理模块；(2) NSF-SRL-NRM $(\alpha = 1/2, \beta = 1, \gamma = 1)$ ： 将视觉推理模块减半；(3) NSF-SRL-OI $(\alpha = 1, \beta = 1, \gamma = 0)$ ： 省略观测变量的交叉熵。我们在数字图像加法和零样本图像分类任务上进行了实验，以评估NSF-SRL及其变体的性能。结果分别展示于图11(a)和图11(b)。

In Fig. 11 (a), we observe that the performance of the NSF-SRL-NRM variant is higher compared to its NSF-SRL-SRM counterparts. This indicates that the symbolic reasoning module is crucial in weakly supervised tasks. This is likely due to the limited availability of supervised information in such tasks. Specifically, in weakly supervised tasks, the input images are not individually labeled but only labeled by the addition task. As a result, the NRM module may have a more restricted role in these tasks. Moreover, this finding highlights the importance of incorporating symbolic knowledge.
在图11(a)中，我们观察到NSF-SRL-NRM变体的性能高于NSF-SRL-SRM变体。这表明符号推理模块在弱监督任务中至关重要。这很可能是由于此类任务中监督信息有限。具体来说，在弱监督任务中，输入图像未被单独标注，而仅通过加法任务进行标注。因此，NRM模块在这些任务中的作用可能较为有限。此外，该发现强调了引入符号知识的重要性。

In Fig. 11 (b), we observe that the correlations among the components of SRM, VRM, and OI have a significantly positive impact on zero-shot image classification. Furthermore, the performance of our model is notably enhanced when SRM is applied, confirming the effectiveness of the symbolic knowledge integrated into the model. We conclude that symbolic knowledge helps the model adapt to new environments, specifically in recognizing unseen classes.
在图11(b)中，我们观察到SRM、VRM和OI组件之间的相关性对零样本图像分类具有显著正面影响。此外，当应用SRM时，我们模型的性能显著提升，验证了模型中集成的符号知识的有效性。我们得出结论，符号知识有助于模型适应新环境，特别是在识别未见类别时。

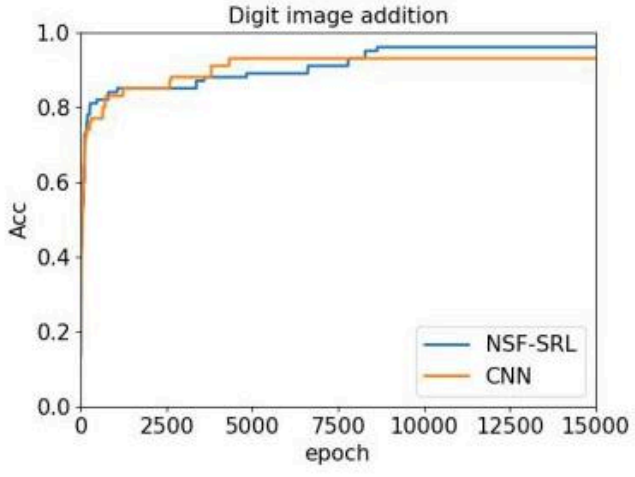
10.15 5.8 Hyperparameter Analysis

10.16 5.8 超参数分析

To analyze the robustness of our NSF-SRL framework and determine optimal hyperparameters, we conducted extensive experiments to evaluate the effects of epoch settings and loss weights (in Eq. 9)).
为了分析NSF-SRL框架的鲁棒性并确定最优超参数，我们进行了大量实验，评估了训练轮数(epoch)设置和损失权重(见公式9)的影响。

1. Effects of Epoch: In Fig. 12, we present the fine-tuning results for models trained with varying numbers epochs, evaluated based on accuracy (Acc) for both digit image addition and zero-shot image classification tasks. The figures clearly show that both NSF-SRL and the baseline models exhibit an upward trend as the number of iterations increases. This trend suggests that the models continue to benefit from longer training, indicating that extended training can further improve performance until convergence. Additionally, the baseline models converge faster than NSF-SRL, which may be due to differences in model architecture, such as CNN or LFGAA having fewer parameters to learn.
2. 训练轮数的影响：在图12中，我们展示了不同训练轮数下模型的微调结果，基于数字图像加法和零样本图像分类任务的准确率（Acc）进行评估。图中清晰地显示，NSF-SRL和基线模型随着迭代次数的增加均呈上升趋势。这表明模型在更长时间的训练中持续受益，说明延长训练时间可以进一步提升性能，直到收敛。此外，基线模型的收敛速度快于NSF-SRL，这可能是由于模型架构的差异，例如CNN或LFGAA参数较少，学习负担较轻。

2. Effects of Loss Weights: In this section, we analyze the impact of the loss weights α, β and γ on their respective loss terms. We experimented with a range of values $\{0, 0.5, 1, 1.5$, $2\}$ for these weights across digit image addition and zero-shot image classification tasks. The results are illustrated in Fig. 13 When $0 < \alpha < 0.5$,all evaluation metrics exhibit an upward trend, while for $\alpha > 0.5$,the performance across all evaluation strategies remains consistent. Additionally, NSF-SRL demonstrates relative insensitivity to β and γ when set to larger values (e.g.,greater than 0.5). Based on these observations,we set α, β ,and γ to 1,1,and 1 , respectively, in our experiments.
3. 损失权重的影响：本节分析了损失权重α, β和γ对各自损失项的影响。我们在数字图像加法和零样本图像分类任务中，尝试了权重取值范围为$\{0, 0.5, 1, 1.5, 2\}$。结果如图13所示，当$0 < \alpha < 0.5$时，所有评估指标均呈上升趋势，而当$\alpha > 0.5$时，各评估策略的性能保持稳定。此外，NSF-SRL对较大值（如大于0.5）的β和γ表现出相对不敏感。基于这些观察，我们在实验中将α, β和γ分别设为1、1和1。



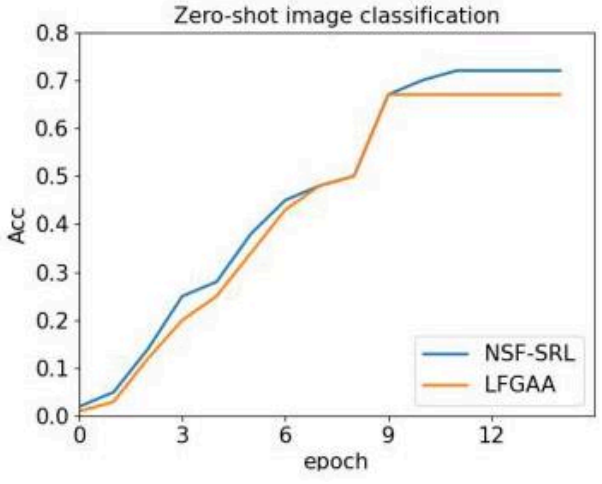


Fig. 12. Effects of different epochs for the NSF-SRL on digit image addition and zero-shot image classification.
图12. 不同训练轮数对NSF-SRL在数字图像加法和零样本图像分类任务中的影响。

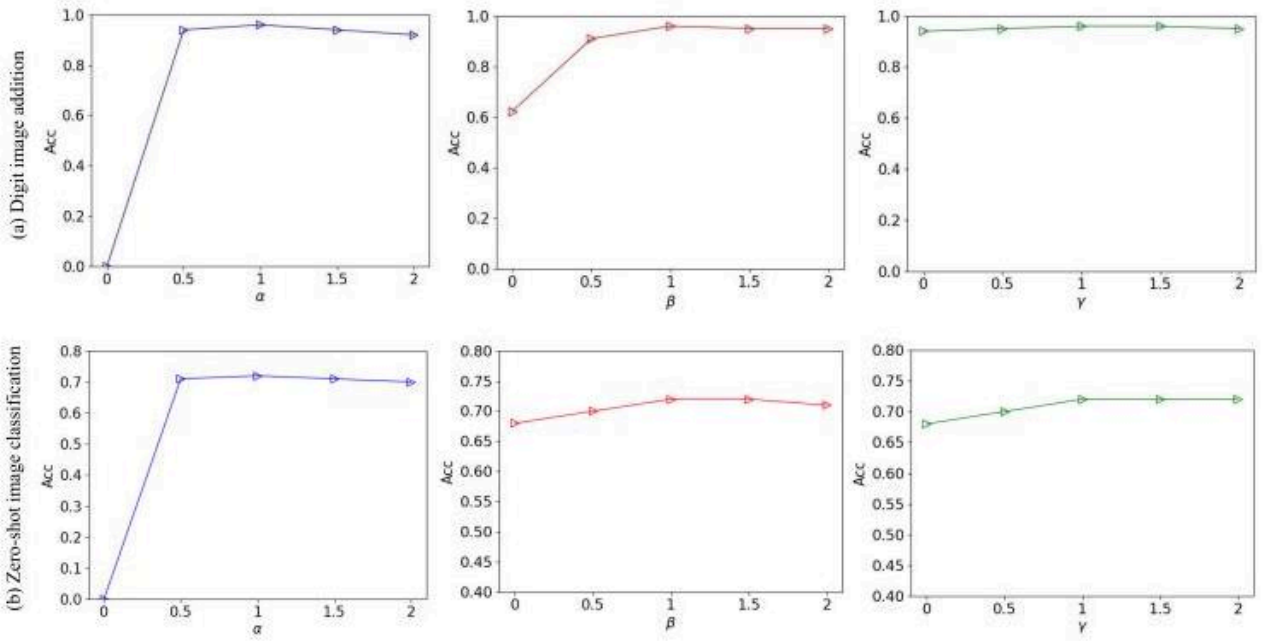


Fig. 13. Effects of loss weights that control their corresponding loss terms on digit image addition and zero-shot image classification tasks, i.e., α, β and γ .
图13. 控制对应损失项的损失权重对数字图像加法和零样本图像分类任务的影响，即α, β和γ。

11 6 CONCLUSION

12 6 结论

In this study, we introduce NSF-SRL, a general model in neural-symbolic systems. Our goal is to improve the model's performance and generalization, while also providing interpretability of the results. Additionally, we propose a novel evaluation metric to quantify the interpretability of the deep model. Our experimental results demonstrate that NSF-SRL outperforms state-of-the-art methods across various reasoning tasks, including supervised, weakly supervised, and zero-shot image classification scenarios, in terms of both performance and generalization. Furthermore, we highlight the interpretability of NSF-SRL by providing visualizations that clarify the model's reasoning process.
本研究提出了NSF-SRL，一种神经符号系统中的通用模型。我们的目标是提升模型的性能和泛化能力，同时提供结果的可解释性。此外，我们提出了一种新颖的评估指标，用以量化深度模型的可解释性。实验结果表明，NSF-SRL在多种推理任务中，包括监督、弱监督和零样本图像分类场景，均优于现有最先进方法，无论是在性能还是泛化能力方面。此外，我们通过可视化展示了NSF-SRL的推理过程，突显其可解释性。

In practice, the NSF-SRL can find applications in diverse scenarios beyond the experimental tasks discussed in this paper. For instance, in healthcare, the model can be leveraged for medical image analysis and patient diagnosis. By amalgamating symbolic reasoning with deep learning capabilities, it can assist physicians in disease diagnosis and treatment planning while enhancing diagnostic reliability through interpretability. In the financial sector, the NSF-SRL can be instrumental in fraud detection and risk assessment by effectively managing complex data patterns with its hybrid approach.

在实际应用中，NSF-SRL可拓展至本文讨论的实验任务之外的多种场景。例如，在医疗领域，该模型可用于医学图像分析和患者诊断。通过结合符号推理与深度学习能力，它能够辅助医生进行疾病诊断和治疗规划，同时通过可解释性提升诊断的可靠性。在金融领域，NSF-SRL可用于欺诈检测和风险评估，凭借其混合方法有效处理复杂数据模式。

In our NSF-SRL framework, the manual definition of logic rules may restrict the breadth of acquired rule knowledge and involves labor costs. On top of this foundational work, a potential enhancement would involve enabling the model to autonomously learn rules from data, leading to a more efficient and adaptive system.
在我们的NSF-SRL框架中，逻辑规则的手动定义可能限制了规则知识的广度，并且涉及人工成本。在此基础上，未来的改进方向是使模型能够自主从数据中学习规则，从而构建更高效且自适应的系统。

13 ACKNOWLEDGMENTS

14 致谢

This work was supported by the National Key R&D Program of China under Grant Nos. 2021ZD0112500; the National Natural Science Foundation of China under Grant Nos. U22A2098, 62172185, 62202200, and 62206105.
REFERENCES
本工作得到中国国家重点研发计划（项目编号2021ZD0112500）和国家自然科学基金（项目编号U22A2098、62172185、62202200、62206105）的资助。参考文献

[1] L. G. Valiant, "Three problems in computer science," in JACM, vol. 50, no. 1, 2003, pp. 96-99.
[1] L. G. Valiant, "计算机科学中的三个问题," JACM, 第50卷，第1期，2003年，页96-99。

[2] V. Belle, "Symbolic logic meets machine learning: A brief survey in infinite domains," in SUM, 2020, pp. 3-16.
[2] V. Belle, "符号逻辑与机器学习的交汇：无限域的简要综述," SUM, 2020年，页3-16。

[3] P. Hitzler and M. K. Sarker, "Neuro-symbolic artificial intelligence: The state of the art," in Neuro-Symbolic Artificial Intelligence, 2021.
[3] P. Hitzler 和 M. K. Sarker, "神经符号人工智能：现状综述," 载于《神经符号人工智能》，2021年。

[4] E. Curry, D. Salwala, P. Dhingra, F. A. Pontes, and P. Yadav, "Multimodal event processing: A neural-symbolic paradigm for the internet of multimedia things," IOTJ, vol. 9, no. 15, pp. 13705-13 724, 2022.
[4] E. Curry, D. Salwala, P. Dhingra, F. A. Pontes 和 P. Yadav, "多模态事件处理：面向多媒体物联网的神经符号范式," IOTJ，第9卷，第15期，2022年，页13705-13724。

[5] D. Yu, B. Yang, D. Liu, H. Wang, and S. Pan, "A survey on neural-symbolic systems," NN, 2022.
[5] D. Yu, B. Yang, D. Liu, H. Wang 和 S. Pan, "神经符号系统综述," NN, 2022年。

[6] M. Qu and J. Tang, "Probabilistic logic neural networks for reasoning," NeurIPS, vol. 32, 2019.
[6] M. Qu 和 J. Tang, “用于推理的概率逻辑神经网络,” NeurIPS, 第32卷, 2019年.

[7] Y. Zhang, X. Chen, Y. Yang, A. Ramamurthy, B. Li, Y. Qi, and L. Song, "Efficient probabilistic logic reasoning with graph neural networks," ICLR, 2020.
[7] Y. Zhang, X. Chen, Y. Yang, A. Ramamurthy, B. Li, Y. Qi 和 L. Song, “基于图神经网络的高效概率逻辑推理,” ICLR, 2020年.

[8] J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu, "The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision," in arXiv preprint arXiv:1904.12584, 2019.
[8] J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum 和 J. Wu, “神经符号概念学习器：通过自然监督解释场景、词汇和句子,” arXiv预印本 arXiv:1904.12584, 2019年.

[9] J. Xu, Z. Zhang, T. Friedman, Y. Liang, and G. Broeck, "A semantic loss function for deep learning with symbolic knowledge," in ICML, 2018, pp. 5502-5511.
[9] J. Xu, Z. Zhang, T. Friedman, Y. Liang 和 G. Broeck, “结合符号知识的深度学习语义损失函数,” ICML, 2018年, 第5502-5511页.

[10] Y. Xie, Z. Xu, M. S. Kankanhalli, K. S. Meel, and H. Soh, "Embedding symbolic knowledge into deep networks," NeurIPS, 2019.
[10] Y. Xie, Z. Xu, M. S. Kankanhalli, K. S. Meel 和 H. Soh, “将符号知识嵌入深度网络,” NeurIPS, 2019年.

[11] R. Luo, N. Zhang, B. Han, and L. Yang, "Context-aware zero-shot recognition," in AAAI, vol. 34, no. 07, 2020, pp. 11709-11716.
[11] R. Luo, N. Zhang, B. Han 和 L. Yang, “上下文感知的零样本识别,” AAAI, 第34卷, 第07期, 2020年, 第11709-11716页.

[12] R. Manhaeve, S. Dumančić, A. Kimmig, T. Demeester, and L. De Raedt, "Neural probabilistic logic programming in deepproblog," AI, vol. 298, p. 103504, 2021.
[12] R. Manhaeve, S. Dumančić, A. Kimmig, T. Demeester 和 L. De Raedt, “DeepProbLog中的神经概率逻辑编程,” AI, 第298卷, 页码103504, 2021年.

[13] Z.-H. Zhou, "Abductive learning: towards bridging machine learning and logical reasoning," SCIS, vol. 62, no. 7, pp. 1-3, 2019.
[13] Z.-H. Zhou, “溯因学习：迈向机器学习与逻辑推理的桥梁,” SCIS, 第62卷, 第7期, 第1-3页, 2019年.

[14] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, and L. De Raedt, "Deepproblog: Neural probabilistic logic programming," NeurIPS, vol. 31, 2018.
[14] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester 和 L. De Raedt, “DeepProbLog：神经概率逻辑编程,” NeurIPS, 第31卷, 2018年.

[15] L. Getoor and B. Taskar, Introduction to statistical relational learning, 2007.
[15] L. Getoor 和 B. Taskar, 统计关系学习导论, 2007年.

[16] D. Yu, B. Yang, Q. Wei, A. Li, and S. Pan, "A probabilistic graphical model based on neural-symbolic reasoning for visual relationship detection," in CVPR, 2022, pp. 10609-10618.
[16] D. Yu, B. Yang, Q. Wei, A. Li 和 S. Pan, “基于神经符号推理的概率图模型用于视觉关系检测,” CVPR, 2022年, 第10609-10618页.

[17] R. Abboud, I. Ceylan, and T. Lukasiewicz, "Learning to reason: Leveraging neural networks for approximate dnf counting," in AAAI, vol. 34, no. 04, 2020, pp. 3097-3104.
[17] R. Abboud, I. Ceylan 和 T. Lukasiewicz, “学习推理：利用神经网络进行近似DNF计数,” AAAI, 第34卷, 第04期, 2020年, 第3097-3104页.

[18] G. Marra and O. Kuželka, "Neural markov logic networks," in Uncertainty in Artificial Intelligence, 2021, pp. 908-917.
[18] G. Marra 和 O. Kuželka, “神经马尔可夫逻辑网络,” 不确定性人工智能会议, 2021年, 第908-917页.

[19] Z. Hu, X. Ma, Z. Liu, E. Hovy, and E. Xing, "Harnessing deep neural networks with logic rules," $ACL, 2016$.
[19] Z. Hu, X. Ma, Z. Liu, E. Hovy 和 E. Xing, “结合逻辑规则利用深度神经网络,” $ACL, 2016$.

[20] Y. Sun, D. Tang, N. Duan, Y. Gong, X. Feng, B. Qin, and D. Jiang, "Neural semantic parsing in low-resource settings with back-translation and meta-learning," in AAAI, vol. 34, no. 05, 2020, pp. 8960-8967.
[20] Y. Sun, D. Tang, N. Duan, Y. Gong, X. Feng, B. Qin 和 D. Jiang, “低资源环境下结合反向翻译和元学习的神经语义解析,” AAAI, 第34卷, 第05期, 2020年, 第8960-8967页.

[21] A. Oltramari, J. Francis, F. Ilievski, K. Ma, and R. Mirzaee, "Generalizable neuro-symbolic systems for commonsense question answering," in Neuro-Symbolic Artificial Intelligence: The State of the Art, 2021, pp. 294-310.
[21] A. Oltramari, J. Francis, F. Ilievski, K. Ma 和 R. Mirzaee, “面向常识问答的可泛化神经符号系统,” 《神经符号人工智能：技术现状》, 2021年, 第294-310页.

[22] S. Badreddine, A. d. Garcez, L. Serafini, and M. Spranger, "Logic tensor networks," vol. 303, p. 103649, 2022.
[22] S. Badreddine, A. d. Garcez, L. Serafini, 和 M. Spranger, “逻辑张量网络（Logic Tensor Networks）,” 卷303, 页103649, 2022年。

[23] J. Tian, Y. Li, W. Chen, L. Xiao, H. He, and Y. Jin, "Weakly supervised neural symbolic learning for cognitive tasks," AAAI, 2022.
[23] J. Tian, Y. Li, W. Chen, L. Xiao, H. He, 和 Y. Jin, “认知任务的弱监督神经符号学习,” AAAI, 2022年。

[24] X. Duan, X. Wang, P. Zhao, G. Shen, and W. Zhu, "Deeplogic: Joint learning of neural perception and logical reasoning," TPAMI, 2022.
[24] X. Duan, X. Wang, P. Zhao, G. Shen, 和 W. Zhu, “Deeplogic：神经感知与逻辑推理的联合学习,” TPAMI, 2022年。

[25] C. Pryor, C. Dickens, E. Augustine, A. Albalak, W. Y. Wang, and L. Getoor, "Neupsl: neural probabilistic soft logic," in IJCAI, 2023, pp. 4145-4153.
[25] C. Pryor, C. Dickens, E. Augustine, A. Albalak, W. Y. Wang, 和 L. Getoor, “Neupsl：神经概率软逻辑,” 见IJCAI, 2023年, 页4145-4153。

[26] Z. Yang, A. Ishay, and J. Lee, "Neurasp: embracing neural networks into answer set programming," in IJCAI, 2021, pp. 1755-1762.
[26] Z. Yang, A. Ishay, 和 J. Lee, “Neurasp：将神经网络融入答案集编程,” 见IJCAI, 2021年, 页1755-1762。

[27] S. D. Tran and L. S. Davis, "Event modeling and recognition using markov logic networks," in ECCV,2008,pp. 610-623.
[27] S. D. Tran 和 L. S. Davis, “基于马尔可夫逻辑网络的事件建模与识别,” 见ECCV, 2008年, 页610-623。

[28] H. Poon and P. Domingos, "Unsupervised semantic parsing," in EMNLP, 2009, pp. 1-10.
[28] H. Poon 和 P. Domingos, “无监督语义解析,” 见EMNLP, 2009年, 页1-10。

[29] W. Zhang, X. Li, H. He, and X. Wang, "Identifying network public opinion leaders based on markov logic networks," The scientific world journal, vol. 2014, 2014.
[29] W. Zhang, X. Li, H. He, 和 X. Wang, “基于马尔可夫逻辑网络识别网络舆论领袖,” 科学世界杂志, 卷2014, 2014年。

[30] P. Singla and P. Domingos, "Discriminative training of markov logic networks," in AAAI, 2005, pp. 868-873.
[30] P. Singla 和 P. Domingos, “马尔可夫逻辑网络的判别式训练,” 见AAAI, 2005年, 页868-873。

[31] L. Mihalkova and R. J. Mooney, "Bottom-up learning of markov logic network structure," in $ML, 2007$,pp. 625-632.
[31] L. Mihalkova 和 R. J. Mooney, “马尔可夫逻辑网络结构的自底向上学习,” 见$ML, 2007$, 页625-632。

[32] P. Singla and P. Domingos, "Memory-efficient inference in relational domains," in AAAI, 2006, pp. 488-493.
[32] P. Singla 和 P. Domingos, “关系域中的内存高效推理,” 见AAAI, 2006年, 页488-493。

[33] T. Khot, S. Natarajan, K. Kersting, and J. Shavlik, "Learning markov logic networks via functional gradient boosting," in ICDM, 2011, pp. 320-329.
[33] T. Khot, S. Natarajan, K. Kersting, 和 J. Shavlik, “通过函数梯度提升学习马尔可夫逻辑网络,” 见ICDM, 2011年, 页320-329。

[34] S. H. Bach, M. Broecheler, B. Huang, and L. Getoor, "Hinge-loss markov random fields and probabilistic soft logic," JLMR, 2017.
[34] S. H. Bach, M. Broecheler, B. Huang, 和 L. Getoor, “铰链损失马尔可夫随机场与概率软逻辑,” JLMR, 2017年。

[35] H. B. Enderton, A mathematical introduction to logic, 2001.
[35] H. B. Enderton, 《逻辑的数学导论》, 2001年。

[36] M. Richardson and P. Domingos, "Markov logic networks," ML, vol. 62, no. 1, pp. 107-136, 2006.
[36] M. Richardson 和 P. Domingos, “马尔可夫逻辑网络,” ML, 卷62, 期1, 页107-136, 2006年。

[37] V. Novák, I. Perfllieva, and J. Mockor, Mathematical principles of fuzzy logic, 2012, vol. 517.
[37] V. Novák, I. Perfllieva, 和 J. Mockor, 《模糊逻辑的数学原理》, 2012年, 卷517。

[38] C. Lu, R. Krishna, M. Bernstein, and L. Fei-Fei, "Visual relationship detection with language priors," in ECCV, 2016, pp. 852-869.
[38] C. Lu, R. Krishna, M. Bernstein, 和 L. Fei-Fei, "基于语言先验的视觉关系检测," 载于 ECCV, 2016, 页码 852-869。

[39] D. Xu, Y. Zhu, C. B. Choy, and L. Fei-Fei, "Scene graph generation by iterative message passing," in CVPR, 2017, pp. 5410-5419.
[39] D. Xu, Y. Zhu, C. B. Choy, 和 L. Fei-Fei, "通过迭代消息传递生成场景图," 载于 CVPR, 2017, 页码 5410-5419。

[40] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-shot learning-a comprehensive evaluation of the good, the bad and the ugly," in TPAMI, vol. 41, no. 9, 2019, pp. 2251-2265.
[40] Y. Xian, C. H. Lampert, B. Schiele, 和 Z. Akata, "零样本学习——对优点、缺点及不足的全面评估," 载于 TPAMI, 第41卷，第9期，2019年，页码 2251-2265。

[41] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona, "Caltech-ucsd birds 200," 2010.
[41] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, 和 P. Perona, "Caltech-ucsd 鸟类200数据集," 2010年。

[42] J. Zhang, Y. Kalantidis, M. Rohrbach, M. Paluri, A. Elgammal, and M. Elhoseiny, "Large-scale visual relationship understanding," in AAAI, vol. 33, no. 01, 2019, pp. 9185-9194.
[42] J. Zhang, Y. Kalantidis, M. Rohrbach, M. Paluri, A. Elgammal, 和 M. Elhoseiny, "大规模视觉关系理解," 载于 AAAI, 第33卷，第01期，2019年，页码 9185-9194。

[43] S. Dasaratha, S. A. Puranam, K. S. Phogat, S. R. Tiyyagura, and N. P. Duffy, "Deeppsl: end-to-end perception and reasoning," in IJCAI, 2023, pp. 3606-3614.
[43] S. Dasaratha, S. A. Puranam, K. S. Phogat, S. R. Tiyyagura, 和 N. P. Duffy, "DeepPSL：端到端的感知与推理," 载于 IJCAI, 2023年，页码 3606-3614。

[44] A. Paul, N. C. Krishnan, and P. Munjal, "Semantically aligned bias reducing zero shot learning," in CVPR, 2019, pp. 7056-7065.
[44] A. Paul, N. C. Krishnan, 和 P. Munjal, "语义对齐的偏差减少零样本学习," 载于 CVPR, 2019年，页码 7056-7065。

[45] Z. Ding and H. Liu, "Marginalized latent semantic encoder for zero-shot learning," in CVPR, 2019, pp. 6191-6199.
[45] Z. Ding 和 H. Liu, "边缘化潜语义编码器用于零样本学习," 载于 CVPR, 2019年，页码 6191-6199。

[46] G.-S. Xie, L. Liu, X. Jin, F. Zhu, Z. Zhang, J. Qin, Y. Yao, and L. Shao, "Attentive region embedding network for zero-shot learning," in CVPR, 2019, pp. 9384-9393.
[46] G.-S. Xie, L. Liu, X. Jin, F. Zhu, Z. Zhang, J. Qin, Y. Yao, 和 L. Shao, "用于零样本学习的注意力区域嵌入网络," 载于 CVPR, 2019年，页码 9384-9393。

[47] Y. Liu, J. Guo, D. Cai, and X. He, "Attribute attention for semantic disambiguation in zero-shot learning," in ECCV, 2019, pp. 6698-6707.
[47] Y. Liu, J. Guo, D. Cai, 和 X. He, "零样本学习中用于语义消歧的属性注意力," 载于 ECCV, 2019年，页码 6698-6707。

[48] D. Huynh and E. Elhamifar, "Fine-grained generalized zero-shot learning via dense attribute-based attention," in CVPR, 2020, pp. 4483-4493.
[48] D. Huynh 和 E. Elhamifar, "通过基于密集属性的注意力实现细粒度广义零样本学习," 载于 CVPR, 2020年，页码 4483-4493。

[49] W. Xu, Y. Xian, J. Wang, B. Schiele, and Z. Akata, "Attribute prototype network for zero-shot learning," NeurIPS, vol. 33, pp. 21969-21980, 2020.
[49] W. Xu, Y. Xian, J. Wang, B. Schiele, 和 Z. Akata, "零样本学习的属性原型网络," NeurIPS, 第33卷，页码 21969-21980，2020年。

[50] B. Yang, Y. Zhang, Y. Peng, c. Zhang, and J. Hang, "Collaborative filtering based zero-shot learning," Journal of Software, vol. 32, no. 9, pp. 2801-2815, 2021.
[50] B. Yang, Y. Zhang, Y. Peng, C. Zhang, 和 J. Hang, "基于协同过滤的零样本学习," 软件学报, 第32卷，第9期，页码 2801-2815，2021年。

[51] Z. Chen, Y. Huang, J. Chen, Y. Geng, W. Zhang, Y. Fang, J. Z. Pan, and H. Chen, "Duet: Cross-modal semantic grounding for contrastive zero-shot learning," in AAAI, vol. 37, no. 1, 2023, pp. 405-413.
[51] Z. Chen, Y. Huang, J. Chen, Y. Geng, W. Zhang, Y. Fang, J. Z. Pan, 和 H. Chen, "Duet：用于对比零样本学习的跨模态语义定位," 载于 AAAI, 第37卷，第1期，2023年，页码 405-413。

[52] S. Chen, Z. Hong, G.-S. Xie, W. Yang, Q. Peng, K. Wang, J. Zhao, and X. You, "Msdn: Mutually semantic distillation network for zero-shot learning," in CVPR, 2022, pp. 7612-7621.
[52] S. Chen, Z. Hong, G.-S. Xie, W. Yang, Q. Peng, K. Wang, J. Zhao, 和 X. You, "MSDN：用于零样本学习的互语义蒸馏网络," 载于 CVPR, 2022年，页码 7612-7621。

[53] D. Huynh and E. Elhamifar, "Compositional zero-shot learning via fine-grained dense feature composition," NeurIPS, vol. 33, pp. 19849-19860, 2020.
[53] D. Huynh 和 E. Elhamifar, "通过细粒度密集特征组合实现组合式零样本学习," NeurIPS, 第33卷，页码 19849-19860，2020年。

[54] S. Chen, Z. Hong, Y. Liu, G.-S. Xie, B. Sun, H. Li, Q. Peng, K. Lu, and X. You, "Transzero: Attribute-guided transformer for zero-shot learning," in *AAAI* ,vol. 36,no. 1,2022,pp. 330-338.
[54] S. Chen, Z. Hong, Y. Liu, G.-S. Xie, B. Sun, H. Li, Q. Peng, K. Lu, 和 X. You, “Transzero：基于属性引导的零样本学习变换器”，发表于*AAAI*，第36卷，第1期，2022年，页330-338。

[55] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in ECCV, 2014, pp. 740-755.
[55] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, 和 C. L. Zitnick, “Microsoft COCO：上下文中的常见物体”，发表于ECCV，2014年，页740-755。

[56] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in NeurIPS, 2013, pp. 3111-3119.
[56] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, 和 J. Dean, “词和短语的分布式表示及其组合性”，发表于NeurIPS，2013年，页3111-3119。

[57] R. Yu, A. Li, V. I. Morariu, and L. S. Davis, "Visual relationship detection with internal and external linguistic knowledge distillation," in ECCV, 2017, pp. 1974-1982.
[57] R. Yu, A. Li, V. I. Morariu, 和 L. S. Davis, “结合内部与外部语言知识蒸馏的视觉关系检测”，发表于ECCV，2017年，页1974-1982。

[58] G. Yin, L. Sheng, B. Liu, N. Yu, X. Wang, J. Shao, and C. C. Loy, "Zoom-net: Mining deep feature interactions for visual relationship recognition," in *ECCV*, 2018 ,pp. 322-338.
[58] G. Yin, L. Sheng, B. Liu, N. Yu, X. Wang, J. Shao, 和 C. C. Loy, “Zoom-net：挖掘深度特征交互以实现视觉关系识别”，发表于*ECCV*, 2018，页322-338。

[59] Y. Zhan, J. Yu, T. Yu, and D. Tao, "On exploring undetermined relationships for visual relationship detection," in CVPR, 2019, pp. 5128-5137.
[59] Y. Zhan, J. Yu, T. Yu, 和 D. Tao, “探索不确定关系用于视觉关系检测”，发表于CVPR，2019年，页5128-5137。

[60] X. Lin, C. Ding, J. Zeng, and D. Tao, "Gps-net: Graph property sensing network for scene graph generation," in CVPR, 2020, pp. 3746-3753.
[60] X. Lin, C. Ding, J. Zeng, 和 D. Tao, “GPS-Net：用于场景图生成的图属性感知网络”，发表于CVPR，2020年，页3746-3753。

[61] Z.-S. Hung, A. Mallya, and S. Lazebnik, "Contextual translation embedding for visual relationship detection and scene graph generation," TPAMI, vol. 43, no. 11, pp. 3820-3832, 2020.
[61] Z.-S. Hung, A. Mallya, 和 S. Lazebnik, “用于视觉关系检测和场景图生成的上下文翻译嵌入”，TPAMI，第43卷，第11期，页3820-3832，2020年。

[62] Y. Hu, S. Chen, X. Chen, Y. Zhang, and X. Gu, "Neural message passing for visual relationship detection," arXiv preprint arXiv:2208.04165, 2022.
[62] Y. Hu, S. Chen, X. Chen, Y. Zhang, 和 X. Gu, “用于视觉关系检测的神经消息传递”，arXiv预印本arXiv:2208.04165，2022年。

[63] A. Newell and J. Deng, "Pixels to graphs by associative embedding," NeurIPS, 2017.
[63] A. Newell 和 J. Deng, “通过关联嵌入实现像素到图的转换”，NeurIPS，2017年。

[64] J. Yang, J. Lu, S. Lee, D. Batra, and D. Parikh, "Graph r-cnn for scene graph generation," in *ECCV*, 2018 ,pp. 670-685.
[64] J. Yang, J. Lu, S. Lee, D. Batra, 和 D. Parikh, “用于场景图生成的Graph R-CNN”，发表于*ECCV*, 2018，页670-685。

[65] R. Herzig, M. Raboh, G. Chechik, J. Berant, and A. Globerson, "Mapping images to scene graphs with permutation-invariant structured prediction," NeurIPS, vol. 31, pp. 7211-7221, 2018.
[65] R. Herzig, M. Raboh, G. Chechik, J. Berant, 和 A. Globerson, “利用置换不变结构化预测将图像映射到场景图”，NeurIPS，第31卷，页7211-7221，2018年。

[66] R. Zellers, M. Yatskar, S. Thomson, and Y. Choi, "Neural motifs: Scene graph parsing with global context," in CVPR, 2018, pp. 5831-5840.
[66] R. Zellers, M. Yatskar, S. Thomson, 和 Y. Choi, “神经模式：基于全局上下文的场景图解析”，发表于CVPR，2018年，页5831-5840。