Learning Algebraic Representation for Systematic Generalization in Abstract Reasoning

Chi Zhang*[1], Sirui Xie*[1], Baoxiong Jia*[1]
Ying Nian Wu[1], Song-Chun Zhu[1,2,3,4], and Yixin Zhu[2]

[1] University of California, Los Angeles, CA 90095, USA
[2] Institute for Artificial Intelligence, Peking University, Beijing 10080, China
[3] Tsinghua University, Beijing 10080, China
[4] Beijing Institute for General Artificial Intelligence, Beijing 10080, China
* indicates equal contribution
{chi.zhang,srxie,baoxiongjia}@ucla.edu, {ywu,sczhu}@stat.ucla.edu, yixin.zhu@pku.edu.cn
Project Website: http://wellyzhang.github.io/project/alans.html

Abstract. Is intelligence realized by connectionist or classicist? While connectionist approaches have achieved superhuman performance, there has been growing evidence that such task-specific superiority is particularly fragile in *systematic generalization*. This observation lies in the central debate between connectionist and classicist, wherein the latter continually advocates an *algebraic* treatment in cognitive architectures. In this work, we follow the classicist's call and propose a hybrid approach to improve systematic generalization in reasoning. Specifically, we showcase a prototype with algebraic representation for the abstract spatial-temporal reasoning task of Raven's Progressive Matrices (RPM) and present the ALgebra-Aware Neuro-Semi-Symbolic (ALANS) learner. The ALANS learner is motivated by abstract algebra and the representation theory. It consists of a neural visual perception frontend and an algebraic abstract reasoning backend: the frontend summarizes the visual information from object-based representation, while the backend transforms it into an algebraic structure and induces the hidden operator on the fly. The induced operator is later executed to predict the answer's representation, and the choice most similar to the prediction is selected as the solution. Extensive experiments show that by incorporating an algebraic treatment, the ALANS learner outperforms various pure connectionist models in domains requiring systematic generalization. We further show the generative nature of the learned algebraic representation; it can be decoded by isomorphism to generate an answer.

1 Introduction

"Thought is in fact a kind of Algebra."

—William James [25]

Imagine you are given two alphabetical sequences of "c, b, a" and "d, c, b", and asked to fill in the missing element in "$e, d, ?$". In nearly no time will one realize

the answer to be c. However, more surprising for human learning is that, effortlessly and instantaneously, we can "freely generalize" [37] the solution to any partial consecutive ordered sequences. While believed to be innate in early development for human infants [39], such systematic generalizability has constantly been missing and proven to be particularly challenging in existing connectionist models [2,29]. In fact, such an ability to entertain a given thought and semantically related contents strongly implies an abstract algebra-like treatment [12]; in literature, it is referred to as the "language of thought" [11], "physical symbol system" [44], and "algebraic mind" [37]. However, in stark contrast, existing connectionist models tend only to capture statistical correlation [7,26,29], rather than providing any account for a structural inductive bias where systematic algebra can be carried out to facilitate generalization.

This contrast instinctively raises a question—what constitutes such an *algebraic* inductive bias? We argue that the foundation of modeling counterpart to the algebraic treatment in early human development [37,39] lies in algebraic computations set up on mathematical axioms, a form of formalized human intuition and the beginning of modern mathematical reasoning [17,34]. Of particular importance to the algebra's basic building blocks is the Peano Axiom [46]. In the Peano Axiom, the essential components of algebra, the algebraic set, and corresponding operators over it, are governed by three statements: (1) the existence of at least one element in the field to study ("zero" element), (2) a successor function that is recursively applied to all elements and can, therefore, span the entire field, and (3) the principle of mathematical induction. Building on such a fundamental axiom, we begin to form the notion of an algebraic set and induce the operator to construct an algebraic structure. We hypothesize that such an algebraic treatment set up on fundamental axioms is essential for a model's systematic generalizability, the lack of which will only make it sub-optimal.

To demonstrate the benefits of adopting such an algebraic treatment in systematic generalization, we showcase a prototype for Raven's Progressive Matrices (RPM) [48,49], an exemplar task for abstract spatial-temporal reasoning [51,68]. In this task, an agent is given an incomplete 3×3 matrix consisting of eight context panels with the last one missing, and asked to pick one answer from a set of eight choices that best completes the matrix. Human's reasoning capability of solving this abstract reasoning task has been commonly regarded as an indicator of "general intelligence" [4] and "fluid intelligence" [19,24,55,56]. In spite of the task being one that ideally requires abstraction, algebraization, induction, and generalization [4,48,49], recent endeavors unanimously propose pure connectionist models that attempt to circumvent such intrinsic cognitive requirements [21,51,58,64,68,70,74]. However, these methods' inefficiency is also evident in systematic generalization; they struggle to extrapolate to domains beyond training [51,70], shown also in this paper.

To address the issue, we introduce an ALgebra-Aware Neuro-Semi-Symbolic (ALANS) learner. At a high-level, the ALANS learner is embedded in a general neuro-symbolic architecture [14,36,66,67] but has the on-the-fly operator learnability, hence **semi-symbolic**. Specifically, it consists of a neural visual per-

ception frontend and an algebraic abstract reasoning backend. For each RPM instance, the neural visual perception frontend first slides a window over each panel to obtain the object-based representation [26,63] for every object. A belief inference engine latter aggregates all object-based representation in each panel to produce the probabilistic *belief state*. The algebraic abstract reasoning backend then takes the belief states of the eight context panels, treats them as snapshots on an algebraic structure, lifts them into a matrix-based algebraic representation built on the Peano Axiom and the representation theory [23], and induces the hidden operator in the algebraic structure by solving an inner optimization problem [3,8,72]. The answer's algebraic representation is predicted by executing the induced operator: its corresponding set element is decoded by isomorphism, and the final answer is selected as the one most similar to the prediction.

The ALANS learner enjoys several benefits in abstract reasoning with an algebraic treatment:

1. Unlike previous monolithic models, the ALANS learner offers a more **interpretable** account of the entire abstract reasoning process: the neural visual perception frontend extracts object-based representation and produces belief states of panels by explicit probability inference, whereas the algebraic abstract reasoning backend induces the hidden operator in the algebraic structure. The final answer's representation is obtained by executing the induced operator, and the choice panel with minimum distance is selected. This process much resembles the top-down bottom-up strategy in human reasoning missed in recent literature [21,51,58,64,68,70,74]: humans reason by inducing the hidden relation, executing it to generate a feasible solution in mind, and choosing the most similar answer available [4].
2. While keeping the semantic interpretability and end-to-end trainability in existing neuro-symbolic frameworks [14,36,66,67], ALANS is **semi-symbolic** in the sense that the symbolic operator can be learned and concluded on the fly without manual definition for every one of them. Such an inductive ability also enables a greater extent of the desired generalizability.
3. By decoding the predicted representation in the algebraic structure, we can also generate an answer that satisfies the hidden relation in the context.

This work makes three major contributions. (1) We propose the ALANS learner, a neuro-semi-symbolic design, in contrast to existing monolithic models. (2) To demonstrate the efficacy of incorporating an algebraic treatment in reasoning, we show the superior systematic generalization ability of the proposed ALANS learner in various extrapolatory RPM domains. (3) We present analyses into both neural visual perception and algebraic abstract reasoning.

2 Related Work

2.1 Quest for Symbolized Manipulation

The idea to treat thinking as a mental language can be dated back to Augustine [1,62]. Since the 1970s, this school of thought has undergone a dramatic

revival as the quest for symbolized manipulation in cognitive modeling, such as "language of thought" [11], "physical symbol system" [44], and "algebraic mind" [37]. In their study, connectionist's task-specific superiority and inability to generalize beyond training [7,26,51,68] have been hypothetically linked to a lack of such symbolized algebraic manipulation [7,29,38]. With evidence that an algebraic treatment adopted in early human development [39] can potentially address the issue [2,36,38], classicist [12] approaches for generalizable reasoning used in programs [40] and blocks world [61] have resurrected. As a hybrid approach to bridge connectionist and classicist, recent developments lead to neuro-symbolic architectures. In particular, the community of theorem proving has been one of the earliest to endorse the technique [13,50,53]: ∂ILP [10] and NLM [9] make inductive programming end-to-end, and DeepProbLog [35] connects learning and reasoning. Recently, Hudson and Manning [22] propose NSM for visual question answering where a probabilistic graph is used for reasoning. Yi *et al.* [67] demonstrate a neuro-symbolic prototype for the same task where a perception module and a language parsing module are separately trained, with the predefined logic operators associated with language tokens chained to process the visual information. Mao *et al.* [36] soften the predefined operators to afford end-to-end training with only question answers. Han *et al.* [14] use the hybrid architecture for metaconcept learning. Yi *et al.* [66] and Chen *et al.* [6] show how neuro-symbolic models can handle explanatory, predictive, and counterfactual questions in temporal and causal reasoning. Lately, NeSS [5] exemplifies an algorithmic stack machine that can be used to improve generalization in language learning. ALANS follows the classicist's call but adopts a neuro-*semi*-symbolic architecture: it is end-to-end trainable as opposed to Yi *et al.* [66,67] and the operator can be learned and concluded on the fly without manual specification.

2.2 Abstract Visual Reasoning

Recent works by Santoro *et al.* [51] and Zhang *et al.* [68] arouse the community's interest in abstract visual reasoning; the task of Raven's Progressive Matrices (RPM) is introduced as such a measure for intelligent agents. As an intelligence quotient test for humans [48,49], RPM is believed to be strongly correlated with human's general intelligence [4] and fluid intelligence [19,24,55,56]. Early RPM-solving systems employ symbolic representation based on hand-designed features and assume access to the underlying logics [4,31,32,33]. Another stream of research on RPM recruits similarity-based metrics to select the most similar answer from the choices [20,30,41,42,43,54]. However, these visual or semantic features are unable to handle uncertainty from imperfect perception, and directly assuming access to the logic operations simplifies the problem. Recently proposed data-driven approaches arise from the availability of large datasets: Santoro *et al.* [51] extend a pedagogical RPM generation method [59], whereas Zhang *et al.* [68] use a stochastic image grammar [75] and introduce structural annotations in it, which Hu *et al.* [21] further refine to avoid shortcut solutions by statistics in candidate panels. Despite the fact that RPM intrinsically requires one to perform abstraction, algebraization, induction, and generalization,

Fig. 1. An overview of the ALANS learner. For an RPM instance, the neural visual perception module produces the belief states for all panels: an object CNN extracts object attribute distributions for each image region, and a belief inference engine marginalizes them out to obtain panel attribute distributions. For each panel attribute, the algebraic abstract reasoning module transforms the belief states into matrix-based algebraic representation and induces hidden operators by solving inner optimizations. The answer representation is obtained by executing the induced operators, and the choice most similar to the prediction is selected as the solution. An example of the underlying discrete algebra and its correspondence is also shown on the right.

existing methods bypass such cognitive requirements using a single feedforward pass in connectionist models: Santoro *et al.* [51] use a relational module [52], Steenbrugge *et al.* [57] augment it with a VAE [28], Zhang *et al.* [68] assemble a dynamic tree, Hill *et al.* [18] arrange the data in a contrastive manner, Zhang *et al.* [70] propose a contrast module, Zhang *et al.* [74] formulate it in a student-teacher setting, Wang *et al.* [58] build a multiplex graph network, Hu *et al.* [21] aggregate features from a hierarchical decomposition, and Wu *et al.* [64] apply a scattering transformation to learn objects, attributes, and relations. Recently, Zhang *et al.* [71] employ a neuro-symbolic design but requires full knowledge over the hidden relations to perform *abduction*. While our work adopts the visual perception module and employs a similar training strategy from Zhang *et al.* [71], the ALANS learner manages to *induce* the hidden relations, enabling on-the-fly relation induction and systematic generalization on relational learning. The recent work of Neural Interpreter (NI) [47] is a complementary neural approach to our method: Although both NI and ALANS decompose the reasoning process into sub-components and aggregate them, NI focuses more on compositionality, routing new input via different paths of learned modules to generalize, whereas ALANS more on induction, enabling a learned module to adapt on the fly.

3 The ALANS Learner

In this section, we introduce the ALANS learner for the RPM problem. In each RPM instance, an agent is given an incomplete 3×3 panel matrix with the last

entry missing and asked to induce the operator hidden in the matrix and choose from eight choice panels one that follows it. Formally, let the answer variable be denoted as y, the context panels as $\{I_{o,i}\}_{i=1}^{8}$, and choice panels as $\{I_{c,i}\}_{i=1}^{8}$. Then the problem can be formulated as estimating $P(y \mid \{I_{o,i}\}_{i=1}^{8}, \{I_{c,i}\}_{i=1}^{8})$. According to the common design [4,51,68], there is one operator that governs each panel attribute. Hence, by assuming independence among attributes, we propose to factorize the probability of $P(y = n \mid \{I_{o,i}\}_{i=1}^{8}, \{I_{c,i}\}_{i=1}^{8})$ as

$$\prod_{a} \sum_{\mathcal{T}^a} P(y^a = n \mid \mathcal{T}^a, \{I_{o,i}\}_{i=1}^{8}, \{I_{c,i}\}_{i=1}^{8}) \times P(\mathcal{T}^a \mid \{I_{o,i}\}_{i=1}^{8}), \tag{1}$$

where y^a denotes the answer selection based only on attribute a and $\mathcal{T}^a$ the operator on a.

Overview As shown in Fig. 1, the ALANS learner decomposes the process into perception and reasoning: the neural visual perception frontend is adopted from Zhang *et al.* [71] and extracts the *belief states* from each of the sixteen panels, whereas the algebraic abstract reasoning backend views an instance as an example in an abstract algebra structure, transforms belief states into *algebraic representation* by the representation theory, *induces* the hidden operators, and *executes* the operators to predict the representation of the answer. Therefore, in Eq. (1), the operator distribution is modeled by the fitness of an operator and the answer distribution by the distance between the predicted representation and that of a candidate.

3.1 Neural Visual Perception

We follow the design in Zhang *et al.* [71] and decompose visual perception into an object CNN and a belief state inference engine. Specifically, for each panel, we use a sliding window to traverse the spatial domain of the image and feed each image region into an object CNN. The CNN has four branches, producing for each region its object attribute distributions, including objectiveness (if the region contains an object), type, size, and color. The belief inference engine summarizes the panel attribute distributions (over position, number, type, size, and color) by marginalizing out all object attribute distributions (over objectiveness, type, size, and color). As an example, the distribution of the panel attribute of Number can be computed as such: for N image regions and their predicted objectiveness

$$P(\text{Number} = k) = \sum_{\substack{R^o \in \{0,1\}^N \\ \sum_j R_j^o = k}} \prod_{j=1}^{N} P(r_j^o = R_j^o), \tag{2}$$

where $P(r_j^o)$ denotes the jth region's estimated objectiveness distribution, and R^o is a binary sequence of length N that sums to k. All panel attribute distributions compose the *belief state* of a panel. In the following, we denote the belief state as b and the distribution of an attribute a as $P(b^a)$. For more details, please refer to Zhang *et al.* [71] and the Appendix.

3.2 Algebraic Abstract Reasoning

Given the belief states of both context and choice panels, the algebraic abstract reasoning backend concerns the induction of hidden operators and the prediction of answer representation for each attribute. The fitness of induced operators is used for estimating the operator distribution and the difference between the prediction and the choice panel for estimating the answer distribution.

Algebraic Underpinning Without loss of generality, here we assume row-wise operators. For each attribute, under perfect perception, the first two rows in an RPM instance provide snapshots into an example of *group* [15] constrained to an integer-indexed set, a simple algebra structure that is closed under a binary operator. To see this, note that an accurate perception module would see each panel attribute as a deterministic set element. Therefore, RPM instances with unary operators, such as progression, are group examples with special binary operators where one operand is constant. Instances with binary operators, such as arithmetics, directly follow the group properties. Those with ternary operators are ones defined on a three-tuple set from rows.

Algebraic Representation A systematic algebraic view allows us to felicitously recruit ideas in the representation theory [23] to glean the hidden properties in the abstract structures: it makes abstract algebra amenable by reducing it onto linear algebra. Following the same spirit, we propose to lift both the set elements and the hidden operators to a learnable matrix space. To encode the set element, we employ the Peano Axiom [46]. According to the Peano Axiom, an integer-indexed set can be constructed by (1) a zero element ($\mathbf{0}$), (2) a successor function ($S(\cdot)$), and (3) the principle of mathematical induction, such that the kth element is encoded as $S^k(\mathbf{0})$. Specifically, we instantiate the zero element as a learnable matrix M_0 and the successor function as the matrix-matrix product parameterized by M. In an attribute-specific manner, the representation of an attribute taking the kth value is $(M^a)^k M_0^a$. For operators, we consider them to live in a learnable matrix group of a corresponding dimension, such that the action of an operator on a set can be represented as matrix multiplication. Such algebraic representation establishes an isomorphism between the matrix space and the abstract algebraic structure: abstract elements on the algebraic structure have a bijective mapping to/from the matrix space, and inducing the abstract relation can be reduced to solving for a matrix operator. See Fig. 2 for a graphical illustration of the isomorphism.

Fig. 2. Isomorphism between the abstract algebra and the matrix-based representation. In this view, operator induction is now reduced to solving for a matrix.

Operator Induction Operator induction concerns about finding a concrete operator in the abstract algebraic structure. By the property of closure, we formulate it as an inner-level regularized linear regression problem: a binary operator $\mathcal{T}_b^a$ for attribute a in a group minimizes $\ell_b^a(\mathcal{T})$ defined as

$$\ell_b^a(\mathcal{T}) = \sum_i \mathbb{E}\left[\|M(b_{o,i}^a)\mathcal{T}M(b_{o,i+1}^a) - M(b_{o,i+2}^a)\|_F^2\right] + \lambda_b^a\|\mathcal{T}\|_F^2, \tag{3}$$

where under visual uncertainty, we take the expectation w.r.t. the distributions in the belief states of context panels $P(b_{o,i}^a)$ in the first two rows, and denote its algebraic representation as $M(b_{o,i}^a)$. For unary operators, one operand can be treated as constant and absorbed into $\mathcal{T}$. Note that Eq. (3) admits a closed-form solution (please refer to the Appendix for details). Therefore, the operator can be learned and adapted for different instances of binary relations and concluded on the fly. Such a design also simplifies the recent neuro-symbolic approaches, where every single symbol operator needs to be hand-defined [14,36,66,67]. Instead, we only specify an inner-level optimization framework and allow symbolic operators to be quickly induced based on the neural observations, while keeping the semantic interpretability in the neuro-symbolic methods. Therefore, we term such a design semi-symbolic.

The operator probability in Eq. (1) is then modeled by each operator type's fitness, *e.g.*, for binary,

$$P(\mathcal{T}^a = \mathcal{T}_b^a \mid \{I_{o,i}\}_{i=1}^8) \propto \exp(-\ell_b^a(\mathcal{T}_b^a)). \tag{4}$$

Operator Execution To predict the algebraic representation of the answer, we solve another inner-level optimization similar to Eq. (3), but now treating the representation of the answer as a variable:

$$\widehat{M_b^a} = \arg\min_M \ell_b^a(M) = \mathbb{E}[\|M(b_{o,7}^a)\mathcal{T}_b^a M(b_{o,8}^a) - M\|_F^2], \tag{5}$$

where the expectation is taken w.r.t. context panels in the last row. The optimization also admits a closed-form solution (please refer to the Appendix for details), which corresponds to the execution of the induced operator in Eq. (3).

The predicted representation is decoded probabilistically as the predicted belief state of the solution,

$$P(\widehat{b^a} = k \mid \mathcal{T}^a) \propto \exp(-\|\widehat{M^a} - (M^a)^k M_0^a\|_F^2). \tag{6}$$

Answer Selection Based on Eqs. (1) and (4), estimating the answer distribution is now boiled down to estimating the conditional answer distributions for each attribute. Here, we propose to model it based on the Jensen–Shannon Divergence (JSD) of the predicted belief state and that of a choice,

$$P(y^a = n \mid \mathcal{T}^a, \{I_{o,i}\}_{i=1}^8, \{I_{c,i}\}_{i=1}^8) \propto \exp(-d_n^a), \tag{7}$$

where we define d_n^a as

$$d_n^a = \mathbb{D}_{\mathrm{JSD}}(P(\widehat{b^a} \mid \mathcal{T}^a)\|P(b_{c,n}^a))). \tag{8}$$

Discussion Comparing with the possible problem-solving process by humans [4], we argue that the proposed algebraic abstract reasoning module offers a computational and interpretable counterpart to human-like reasoning in RPM. Specifically, the induction component resembles fluid intelligence, where one quickly induces the hidden operator by observing the context panels. The execution component synthesizes an image by executing the induced operator, and the choice most similar to the image is selected as the answer.

We also note that by decoding the predicted representation in Eq. (6), a solution can be *generated*: by sequentially selecting the most probable operator and the most probable attribute value, a rendering engine can directly render the solution. The reasoning backend also enables end-to-end training: by integrating the belief states from neural perception, the module conducts both induction and execution in a soft manner, such that the gradients can be back-propagated and both the visual frontend and the reasoning backend jointly trained.

3.3 Training Strategy

We train the entire ALANS learner by minimizing the cross-entropy loss between the estimated answer distribution and the ground-truth selection and an auxiliary loss [51,58,68,71] that shapes the operator distribution from the reasoning engine, *i.e.*,

$$\min_{\theta,\{M_0^a\},\{M^a\}} \ell(P(y \mid \{I_{o,i}\}_{i=1}^8, \{I_{c,i}\}_{i=1}^8), y_\star) + \sum_a \lambda^a \ell(P(\mathcal{T}^a \mid \{I_{o,i}\}_{i=1}^8), y_\star^a), \quad (9)$$

where $\ell(\cdot)$ denotes the cross-entropy loss, $y_\star$ the correct choice in candidates, and $y_\star^a$ the ground-truth operator selection for attribute a. The first part of the loss encourages the model to select the right choice for evaluation, while the second part motivates meaningful internal representation to emerge. Compared to Zhang *et al.* [71], the system requires joint operation from not only a trained perception module θ, but also the algebraic encodings from the zero elements $\{M_0^a\}$ and the successor functions $\{M^a\}$, and correspondingly, induced operators $\mathcal{T}$. We notice the three-stage curriculum in Zhang *et al.* [71] is crucial for such a neuro-semi-symbolic system. In particular, we use λ^a to balance the trade-off in the curriculum: in the first stage, we only train parameters regarding objectiveness; in the second stage, we freeze objectiveness parameters and cyclically train parameters involving type, size, and color; in the last stage, we fine-tune all parameters.

4 Experiments

A cognitive architecture with systematic generalization is believed to demonstrate the following three principles [12,37,38]: (1) systematicity, (2) productivity, and (3) localism. Systematicity requires an architecture to be able to entertain "semantically related" contents after understanding a given thought. Productivity states the awareness of a constituent implies that of a recursive application of the constituent; vice versa for localism.

To verify the effectiveness of an algebraic treatment in systematic generalization, we showcase the superiority of the proposed ALANS learner on the three principles in the abstract spatial-temporal reasoning task of RPM. Specifically, we use the generation methods proposed in Zhang *et al.* [68] and Hu *et al.* [21] to generate RPM problems and carefully split training and testing to construct the three regimes. The former generates candidates by perturbing only one attribute of the correct answer while the later modifies attribute values in a hierarchical manner to avoid shortcut solutions by pure statistics. Both methods categorize relations in RPM into three types, according to Carpenter *et al.* [4]: unary (Constant and Progression), binary (Arithmetic), and ternary (Distribution of Three), each of which comes with several instances. Grounding the principles into learning abstract relations in RPM, we fix the configuration to be 3×3Grid and generate the following data splits for evaluation (please refer to the Appendix for details):

- Systematicity: the training set contains only a subset of instances for each type of relation, while the test set all other relation instances.
- Productivity: as the binary relation results from a recursive application of the unary relation, the training set contains only unary relations, whereas the test set only binary relations.
- Localism: the training and testing sets in the productivity split are swapped to study localism.

We follow Zhang *et al.* [68] to generate $10,000$ instances for each split and assign 6 folds for training, 2 folds for validation, and 2 folds for testing.

4.1 Experimental Setup

We evaluate the systematic generalizability of the proposed ALANS learner on the above three splits, and compare the ALANS learner with other baselines, including ResNet [16], ResNet+DRT [68], WReN [51], CoPINet [70], MXGNet [58], LEN [74], HriNet [21], and SCL [64]. We use either official or public implementations that reproduce the original results. All models are implemented in PyTorch [45] and optimized using ADAM [27] on an Nvidia Titan Xp GPU. We validate trained models on validation sets and report performance on test sets.

4.2 Systematic Generalization

Tab. 1 shows the performance of various models on systematic generalization, *i.e.*, systematicity, productivity, and localism. Compared to results reported in existing works mentioned above, all pure connectionist models experience a devastating performance drop when it comes to the critical cognitive requirements on systematic generalization, indicating that pure connectionist models fail to perform abstraction, algebraization, induction, or generalization needed in solving the abstract reasoning task; instead, they seem to only take a shortcut to bypass them. In particular, MXGNet's [58] superiority is diminishing

Table 1. Model performance on different aspects of systematic generalization. The performance is measured by accuracy on the test sets. Results on datasets generated by Zhang *et al.* [68] (upper) and by Hu *et al.* [21] (lower).

Method	MXGNet	ResNet+DRT	ResNet	HriNet	LEN	WReN	SCL	CoPINet	ALANS	ALANS-Ind	ALANS-V
Systematicity	20.95%	33.00%	27.35%	28.05%	40.15%	35.20%	37.35%	59.30%	**78.45**%	52.70%	93.85%
Productivity	30.40%	27.95%	27.05%	31.45%	42.30%	56.95%	51.10%	60.00%	**79.95**%	36.45%	90.20%
Localism	28.80%	24.90%	23.05%	29.70%	39.65%	38.70%	47.75%	60.10%	**80.50**%	59.80%	95.30%
Average	26.72%	28.62%	25.82%	29.73%	40.70%	43.62%	45.40%	59.80%	**79.63**%	48.65%	93.12%
Systematicity	13.35%	13.50%	14.20%	21.00%	17.40%	15.00%	24.90%	18.35%	**64.80**%	52.80%	84.85%
Productivity	14.10%	16.10%	20.70%	20.35%	19.70%	17.95%	22.20%	29.10%	**65.55**%	32.10%	86.55%
Localism	15.80%	13.85%	17.45%	24.60%	20.15%	19.70%	29.95%	31.85%	**65.90**%	50.70%	90.95%
Average	14.42%	14.48%	17.45%	21.98%	19.08%	17.55%	25.68%	26.43%	**65.42**%	45.20%	87.45%

in systematic generalization. In spite of learning with structural annotations, ResNet+DRT [68] does not fare better than its base model. The recently proposed HriNet [21] slightly improves on ResNet [16] in this aspect, with LEN [74] being only marginally better. WReN [51], on the other hand, shows oscillating performance across three regimes. Evaluated under systematic generation, SCL [64] and CoPINet [70] also far deviate from "superior performance." These observations suggest that pure connectionist models highly likely learn from variation in visual appearance rather than the algebra underlying the problem.

Embedded in a neural-semi-symbolic framework, the proposed ALANS learner improves on systematic generalization by a large margin. With an algebra-aware design, the model is considerably stable across different principles of systematic generalization. The algebraic representation learned in relations of either a constituent or a recursive composition naturally supports productivity and localism, while semi-symbolic inner optimization further allows various instances of an operator type to be induced from the algebraic representation and boosts systematicity. The importance of the algebraic representation is made more significant in the ablation study: ALANS-Ind, with algebraic representation replaced by independent encodings and the algebraic isomorphism broken, shows inferior performance. We also examine the performance of the learner with perfect visual annotations (denoted as ALANS-V) to see how the proposed algebraic reasoning module works: the gap despite of accurate perception indicates space for improvement for the inductive reasoning part of the model. In the next section, we further show that the neuro-semi-symbolic decomposition in ALANS's design enables diagnostic tests into its jointly learned perception module and reasoning module. This design is in stark contrast to black-box models.

4.3 Analysis into Perception and Reasoning

The neural-semi-symbolic design affords analyses into both perception and reasoning. To evaluate the neural perception and the algebraic reasoning modules, we extract region-based object attribute annotations from the datasets [21,68] and categorize all relations into three types, *i.e.*, unary, binary, and ternary.

Tab. 2 shows the perception module's performance on the test sets in the three regimes of systematic generalization. We note that in order for the ALANS learner to achieve the desired results shown in Tab. 1, ALANS learns to construct the concept of objectiveness perfectly. The model also shows fairly accurate prediction on the attributes of type and size. However, on the texture-related concept of color, ALANS fails to develop a reliable notion on it. Despite that, the general prediction accuracy of the perception module is still surprising, considering that the perception module is jointly learned with ground-truth annotations on answer selections. The relatively lower accuracy on color could be attributed to its larger space compared to other attributes.

Table 2. Perception accuracy of the proposed ALANS learner, measured by whether the module can correctly predict an attribute's value. Results on datasets generated by Zhang *et al.* [68] (left) and by Hu *et al.* [21] (right).

Object Attribute	Objectiveness	Type	Size	Color
Systematicity	100.00%	99.95%	94.65%	71.35%
Productivity	100.00%	99.97%	98.04%	77.61%
Localism	100.00%	95.65%	98.56%	80.05%
Average	100.00%	98.52%	97.08%	76.34%

Object Attribute	Objectiveness	Type	Size	Color
Systematicity	100.00%	96.34%	92.36%	63.98%
Productivity	100.00%	94.28%	97.00%	69.89%
Localism	100.00%	95.80%	98.36%	60.35%
Average	100.00%	95.47%	95.91%	64.74%

Tab. 3 lists the reasoning module's performance during testing for the three aspects. Note that on position, the unary operator (shifting) and binary operator (set arithmetics) do not systematically imply each other. Hence, we do not count them as probes into productivity and localism. In general, we notice that the better the perception accuracy on one attribute, the better the performance on reasoning. However, we also note that despite the relatively accurate perception of objectiveness, type, and size, near perfect reasoning is never guaranteed. This deficiency is due to the perception uncertainty handled by expectation in Eq. (3): in spite of correctness when we take $\arg\max$, marginalizing by expectation will unavoidably introduce noise into the reasoning process. Therefore, an ideal reasoning module requires the perception frontend to be not only correct but also certain. Computationally, one can sample from the perception module and optimize Eq. (9) using REINFORCE [60]. However, the credit assignment problem and variance in gradient estimation will further complicate training.

4.4 In-Distribution Performance

To further evaluate how models perform under the regular Independent and Identically Distributed (I.I.D.) setup, we train the models on the original datasets generated by Zhang *et al.* [68] and Hu *et al.* [21] and measure the model accuracy in the test splits. We compare ALANS with published baselines in Tab. 1.

Tab. 4 (left) shows the results on the RAVEN dataset [68]. With the jointly trained vision component, the ALANS learner does not fare better than the best connectionist approaches, making it on par with SCL only. As the dataset is

Table 3. Reasoning accuracy of the proposed ALANS learner, measured by whether the module can correctly predict the type of a relation on an attribute. Results on datasets generated by Zhang *et al.* [68] (left) and by Hu *et al.* [21] (right).

Relation on	Position	Number	Type	Size	Color
Systematicity	72.04%	82.14%	81.50%	80.80%	40.40%
Productivity	-	98.75%	89.50%	72.10%	33.95%
Localism	-	74.70%	44.25%	56.40%	54.20%
Average	72.04%	85.20%	71.75%	69.77%	42.85%

Relation on	Position	Number	Type	Size	Color
Systematicity	69.96%	80.34%	83.50%	80.85%	28.85%
Productivity	-	99.10%	87.95%	68.50%	23.10%
Localism	-	70.55%	36.65%	42.30%	33.20%
Average	69.96%	83.33%	69.37%	63.88%	28.38%

Table 4. Model performance on RAVEN [68] (left) and I-RAVEN [21] (right) under the regular I.I.D. evaluation, measured by accuracy on the test sets.

Method	Acc	Center	2x2Grid	3x3Grid	L-R	U-D	O-IC	O-IG
WReN	34.0%/21.5%	58.4%/24.0%	38.9%/25.0%	37.7%/20.1%	21.6%/19.7%	19.8%/19.9%	38.9%/21.3%	22.6%/20.6%
ResNet	53.4%/18.4%	52.8%/22.6%	41.9%/15.5%	44.3%/18.1%	58.8%/19.0%	60.2%/19.6%	63.2%/17.5%	53.1%/16.6%
ResNet+DRT	59.6%/20.7%	58.1%/24.2%	46.5%/18.2%	50.4%/19.8%	65.8%/22.0%	67.1%/22.1%	69.1%/21.0%	60.1%/18.1%
LEN	71.6%/32.8%	79.1%/44.8%	56.1%/27.9%	60.3%/23.9%	80.5%/34.1%	76.4%/34.4%	79.3%/35.8%	69.9%/28.5%
HriNet	45.1%/60.8%	66.1%/78.2%	40.7%/50.1%	38.0%/42.4%	44.9%/70.1%	43.2%/70.3%	47.2%/68.2%	35.8%/46.3%
MXGNet	84.0%/33.1%	94.3%/40.7%	60.5%/27.9%	64.9%/24.7%	96.6%/35.8%	96.4%/34.5%	94.1%/36.4%	81.3%/31.6%
CoPINet	91.4%/46.1%	95.1%/54.4%	77.5%/36.8%	78.9%/31.9%	**99.1**%/51.9%	**99.7**%/52.5%	**98.5**%/52.2%	91.4%/42.8%
ALANS	74.4%/78.5%	69.1%/72.3%	80.2%/79.5%	75.0%/72.9%	72.2%/79.2%	73.3%/79.6%	76.3%/85.9%	74.9%/79.9%
SCL	74.2%/80.5%	82.8%/84.6%	70.4%/79.4%	64.1%/69.9%	77.6%/82.7%	78.4%/82.6%	84.2%/87.3%	62.2%/77.2%
ALANS-V	**94.4**%/**93.5**%	**98.4**%/**98.9**%	**91.5**%/**85.0**%	**87.0**%/**83.2**%	97.3%/**90.9**%	96.4%/**98.1**%	97.3%/**99.1**%	**93.2**%/**89.5**%

known to have shortcut solutions, neural approaches like MXGNet and CoPINet could potentially find it easier to solve and hence achieve much superior results in this setup. However, ALANS-V, the variant with a perfect perception component reaches a level of much robustness and accuracy, attaining the best results in grid-like layouts, empirically believed to be the hardest in human evaluation [68].

Tab. 4 (right) shows the results on the I-RAVEN dataset [21]. Apart from ALANS-V's realizing the best performance across all models, we also notice the consistency of performance of the proposed method across datasets, with or without the shortcut issues. All other methods show drastically varying performance, particularly for CoPINet and MXGNet, arguably because of the choice generation strategy that effectively prunes easy solution paths via statistics.

In summary, by analyzing the results from Tabs. 1 and 4 together, we notice that ALANS not only attains reasonable performance on the I.I.D. setup but also generalizes systematically.

4.5 Generative Potential

Compared to existing discriminative-only RPM-solving methods, the proposed ALANS learner is unique in its generative potential. As mentioned above, the final panel attribute can be decoded by sequentially selecting the most probable hidden operator and the attribute value. A solution can be generated when equipped with a rendering engine. In Fig. 3, we use the rendering program from Zhang *et al.* [68] to showcase the generative potential in the ALANS learner.

Fig. 3. Examples of RPM instances with the missing entries filled by solutions directly generated by the ALANS learner. Ground-truth relations are also listed. Note the generated results do not look exactly the same as the correct candidate choices due to random rotations during rendering, but they are semantically correct.

5 Conclusion and Limitation

In this work, we propose the ALgebra-Aware Neuro-Semi-Symbolic (ALANS) learner, echoing a normative theory in the connectionist-classicist debate that an algebraic treatment in a cognitive architecture should improve a model's systematic generalization ability. In particular, the ALANS learner employs a neural-semi-symbolic architecture, where the neural visual perception module is responsible for summarizing visual information and the algebraic abstract reasoning module transforms it into algebraic representation with isomorphism established by the Peano Axiom and the representation theory, conducts operator induction, and executes it to arrive at an answer. In three RPM domains reflective of systematic generalization, the proposed ALANS learner shows superior performance compared to other pure connectionist baselines.

The proposed ALANS learner also bears some limitations. For one thing, we make the assumption in our formulation that relations on different attributes are independent and can be factorized. This assumption is not universally correct and could potentially lead to failure in more complex reasoning scenarios when attributes are correlated. For another, we assume a fixed and known space for each attribute in the perception module, while in the real world the space for one attribute could be dynamically changing. In addition, the reasoning module is sensitive to perception uncertainty as has already been discussed in the experimental results. Besides, the gap between perfection and the status quo in reasoning remains to be filled. In this work, we only show how the hidden operator can be induced with regularized linear regression via the representation theory. However, more elaborate differentiable optimization problems can certainly be incorporated for other problems.

With the limitation in mind, we hope that this preliminary study could inspire more research on incorporating algebraic structures into current connectionist models and help address challenging modeling problems [65,69,73,76].

Acknowledgement We thank Prof. Hongjing Lu and colleagues from UCLA for fruitful discussions. We would also like to thank anonymous reviewers for constructive feedback.

References

1. Augustine, S.: The confessions. Clark (1876)
2. Bahdanau, D., Murty, S., Noukhovitch, M., Nguyen, T.H., Vries, H.d., Courville, A.: Systematic generalization: What is required and can it be learned? In: International Conference on Learning Representations (ICLR) (2019)
3. Bard, J.F.: Practical bilevel optimization: algorithms and applications, vol. 30. Springer Science & Business Media (2013)
4. Carpenter, P.A., Just, M.A., Shell, P.: What one intelligence test measures: a theoretical account of the processing in the raven progressive matrices test. Psychological Review **97**(3), 404 (1990)
5. Chen, X., Liang, C., Yu, A.W., Song, D., Zhou, D.: Compositional generalization via neural-symbolic stack machines. In: Advances in Neural Information Processing Systems (2020)
6. Chen, Z., Mao, J., Wu, J., Wong, K.Y.K., Tenenbaum, J.B., Gan, C.: Grounding physical concepts of objects and events through dynamic visual reasoning. In: International Conference on Learning Representations (ICLR) (2020)
7. Chollet, F.: The measure of intelligence. arXiv preprint arXiv:1911.01547 (2019)
8. Colson, B., Marcotte, P., Savard, G.: An overview of bilevel optimization. Annals of Operations Research **153**(1), 235–256 (2007)
9. Dong, H., Mao, J., Lin, T., Wang, C., Li, L., Zhou, D.: Neural logic machines. In: International Conference on Learning Representations (ICLR) (2018)
10. Evans, R., Grefenstette, E.: Learning explanatory rules from noisy data. Journal of Artificial Intelligence Research (JAIR) **61**, 1–64 (2018)
11. Fodor, J.A.: The language of thought, vol. 5. Harvard university press (1975)
12. Fodor, J.A., Pylyshyn, Z.W., et al.: Connectionism and cognitive architecture: A critical analysis. Cognition **28**(1-2), 3–71 (1988)
13. Garcez, A.S.d., Broda, K.B., Gabbay, D.M.: Neural-symbolic learning systems: foundations and applications. Springer Science & Business Media (2012)
14. Han, C., Mao, J., Gan, C., Tenenbaum, J., Wu, J.: Visual concept-metaconcept learning. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)
15. Hausmann, B.A., Ore, O.: Theory of quasi-groups. American Journal of Mathematics **59**(4), 983–1004 (1937)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
17. Heath, T.L., et al.: The thirteen books of Euclid's Elements. Courier Corporation (1956)
18. Hill, F., Santoro, A., Barrett, D.G., Morcos, A.S., Lillicrap, T.: Learning to make analogies by contrasting abstract relational structure. In: International Conference on Learning Representations (ICLR) (2019)
19. Hofstadter, D.R.: Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought. Basic books (1995)
20. Holyoak, K.J., Ichien, N., Lu, H.: From semantic vectors to analogical mapping. Current Directions in Psychological Science p. 09637214221098054 (2022)
21. Hu, S., Ma, Y., Liu, X., Wei, Y., Bai, S.: Hierarchical rule induction network for abstract visual reasoning. arXiv preprint arXiv:2002.06838 (2020)
22. Hudson, D., Manning, C.D.: Learning by abstraction: The neural state machine. In: Advances in Neural Information Processing Systems (2019)
23. Humphreys, J.E.: Introduction to Lie algebras and representation theory, vol. 9. Springer Science & Business Media (2012)

24. Jaeggi, S.M., Buschkuehl, M., Jonides, J., Perrig, W.J.: Improving fluid intelligence with training on working memory. Proceedings of the National Academy of Sciences (PNAS) **105**(19), 6829–6833 (2008)
25. James, W.: The Principles of Psychology. Henry Holt and Company (1891)
26. Kansky, K., Silver, T., Mély, D.A., Eldawy, M., Lázaro-Gredilla, M., Lou, X., Dorfman, N., Sidor, S., Phoenix, S., George, D.: Schema networks: Zero-shot transfer with a generative causal model of intuitive physics. In: International Conference on Machine Learning (ICML) (2017)
27. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2014)
28. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
29. Lake, B., Baroni, M.: Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In: International Conference on Machine Learning (ICML) (2018)
30. Little, D.R., Lewandowsky, S., Griffiths, T.L.: A bayesian model of rule induction in raven's progressive matrices. In: Annual Meeting of the Cognitive Science Society (CogSci) (2012)
31. Lovett, A., Forbus, K.: Modeling visual problem solving as analogical reasoning. Psychological Review **124**(1), 60 (2017)
32. Lovett, A., Forbus, K., Usher, J.: A structure-mapping model of raven's progressive matrices. In: Annual Meeting of the Cognitive Science Society (CogSci) (2010)
33. Lovett, A., Tomai, E., Forbus, K., Usher, J.: Solving geometric analogy problems through two-stage analogical mapping. Cognitive Science **33**(7), 1192–1231 (2009)
34. Maddy, P.: Believing the axioms. i. The Journal of Symbolic Logic **53**(2), 481–511 (1988)
35. Manhaeve, R., Dumancic, S., Kimmig, A., Demeester, T., De Raedt, L.: Deepproblog: Neural probabilistic logic programming. In: Advances in Neural Information Processing Systems (NeurIPS) (2018)
36. Mao, J., Gan, C., Kohli, P., Tenenbaum, J.B., Wu, J.: The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. In: International Conference on Learning Representations (ICLR) (2019)
37. Marcus, G.: The algebraic mind. Cambridge, MA: MIT Press (2001)
38. Marcus, G.: The next decade in ai: four steps towards robust artificial intelligence. arXiv preprint arXiv:2002.06177 (2020)
39. Marcus, G.F., Vijayan, S., Rao, S.B., Vishton, P.M.: Rule learning by seven-month-old infants. Science **283**(5398), 77–80 (1999)
40. McCarthy, J.: Programs with common sense. RLE and MIT computation center (1960)
41. McGreggor, K., Goel, A.: Confident reasoning on raven's progressive matrices tests. In: AAAI Conference on Artificial Intelligence (AAAI) (2014)
42. McGreggor, K., Kunda, M., Goel, A.: Fractals and ravens. Artificial Intelligence **215**, 1–23 (2014)
43. Mekik, C.S., Sun, R., Dai, D.Y.: Similarity-based reasoning, raven's matrices, and general intelligence. In: International Joint Conference on Artificial Intelligence (IJCAI) (2018)
44. Newell, A.: Physical symbol systems. Cognitive Science **4**(2), 135–183 (1980)
45. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in PyTorch. In: NIPS Autodiff Workshop (2017)

46. Peano, G.: Arithmetices principia: Nova methodo exposita. Fratres Bocca (1889)
47. Rahaman, N., Gondal, M.W., Joshi, S., Gehler, P., Bengio, Y., Locatello, F., Schölkopf, B.: Dynamic inference with neural interpreters. Advances in Neural Information Processing Systems (NeurIPS) (2021)
48. Raven, J.C.: Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive. Master's thesis, University of London (1936)
49. Raven, J.C., Court, J.H.: Raven's progressive matrices and vocabulary scales. Oxford pyschologists Press (1998)
50. Rocktäschel, T., Riedel, S.: End-to-end differentiable proving. In: Advances in Neural Information Processing Systems (NeurIPS) (2017)
51. Santoro, A., Hill, F., Barrett, D., Morcos, A., Lillicrap, T.: Measuring abstract reasoning in neural networks. In: International Conference on Machine Learning (ICML) (2018)
52. Santoro, A., Raposo, D., Barrett, D.G., Malinowski, M., Pascanu, R., Battaglia, P., Lillicrap, T.: A simple neural network module for relational reasoning. In: Advances in Neural Information Processing Systems (NeurIPS) (2017)
53. Serafini, L., Garcez, A.d.: Logic tensor networks: Deep learning and logical reasoning from data and knowledge. arXiv preprint arXiv:1606.04422 (2016)
54. Shegheva, S., Goel, A.: The structural affinity method for solving the raven's progressive matrices test for intelligence. In: AAAI Conference on Artificial Intelligence (AAAI) (2018)
55. Spearman, C.: The nature of "intelligence" and the principles of cognition. Macmillan (1923)
56. Spearman, C.: The abilities of man, vol. 6. Macmillan New York (1927)
57. Steenbrugge, X., Leroux, S., Verbelen, T., Dhoedt, B.: Improving generalization for abstract reasoning tasks using disentangled feature representations. arXiv preprint arXiv:1811.04784 (2018)
58. Wang, D., Jamnik, M., Lio, P.: Abstract diagrammatic reasoning with multiplex graph networks. In: International Conference on Learning Representations (ICLR) (2020)
59. Wang, K., Su, Z.: Automatic generation of raven's progressive matrices. In: International Joint Conference on Artificial Intelligence (IJCAI) (2015)
60. Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine Learning **8**(3-4), 229–256 (1992)
61. Winograd, T.: Procedures as a representation for data in a computer program for understanding natural language. Tech. rep., MIT. Cent. Space Res. (1971)
62. Wittgenstein, L.: Philosophical investigations. Philosophische Untersuchungen. Macmillan (1953)
63. Wu, J., Tenenbaum, J.B., Kohli, P.: Neural scene de-rendering. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
64. Wu, Y., Dong, H., Grosse, R., Ba, J.: The scattering compositional learner: Discovering objects, attributes, relationships in analogical reasoning. arXiv preprint arXiv:2007.04212 (2020)
65. Xu, M., Jiang, G., Zhang, C., Zhu, S.C., Zhu, Y.: Est: Evaluating scientific thinking in artificial agents. arXiv preprint arXiv:2206.09203 (2022)
66. Yi, K., Gan, C., Li, Y., Kohli, P., Wu, J., Torralba, A., Tenenbaum, J.: Clevrer: Collision events for video representation and reasoning. In: International Conference on Learning Representations (ICLR) (2020)

67. Yi, K., Wu, J., Gan, C., Torralba, A., Kohli, P., Tenenbaum, J.: Neural-symbolic vqa: Disentangling reasoning from vision and language understanding. In: Advances in Neural Information Processing Systems (NeurIPS) (2018)
68. Zhang, C., Gao, F., Jia, B., Zhu, Y., Zhu, S.C.: Raven: A dataset for relational and analogical visual reasoning. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
69. Zhang, C., Jia, B., Edmonds, M., Zhu, S.C., Zhu, Y.: Acre: Abstract causal reasoning beyond covariation. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
70. Zhang, C., Jia, B., Gao, F., Zhu, Y., Lu, H., Zhu, S.C.: Learning perceptual inference by contrasting. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)
71. Zhang, C., Jia, B., Zhu, S.C., Zhu, Y.: Abstract spatial-temporal reasoning via probabilistic abduction and execution. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
72. Zhang, C., Zhu, Y., Zhu, S.C.: Metastyle: Three-way trade-off among speed, flexibility, and quality in neural style transfer. In: AAAI Conference on Artificial Intelligence (AAAI) (2019)
73. Zhang, W., Zhang, C., Zhu, Y., Zhu, S.C.: Machine number sense: A dataset of visual arithmetic problems for abstract and relational reasoning. In: AAAI Conference on Artificial Intelligence (AAAI) (2020)
74. Zheng, K., Zha, Z.J., Wei, W.: Abstract reasoning with distracting features. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)
75. Zhu, S.C., Mumford, D., et al.: A stochastic grammar of images. Foundations and Trends® in Computer Graphics and Vision **2**(4), 259–362 (2007)
76. Zhu, Y., Gao, T., Fan, L., Huang, S., Edmonds, M., Liu, H., Gao, F., Zhang, C., Qi, S., Wu, Y.N., et al.: Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. Engineering **6**(3), 310–345 (2020)