Logical Vision: One-Shot Meta-Interpretive Learning from Real Images

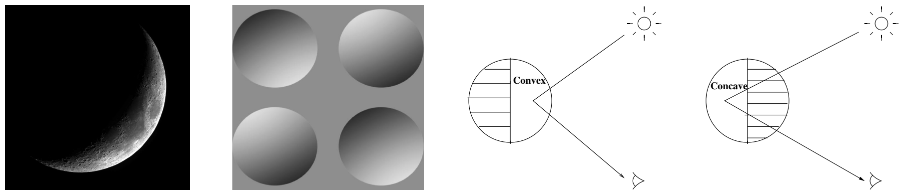
Wang-Zhou Dai[1], Stephen Muggleton[2], Jing Wen[3], Alireza Tamaddoni-Nezhad[2], and Zhi-Hua Zhou[1]

[1] National Key Laboratory for Novel Software Technology, Nanjing University
[2] Department of Computing, Imperial College London
[3] School of Computer and Information Technology, Shanxi University

Abstract. Statistical machine learning is widely used in image classification. However, most techniques 1) require many images to achieve high accuracy and 2) do not provide support for reasoning below the level of classification, and so are unable to support secondary reasoning, such as the existence and position of light sources and other objects outside the image. In recent work an Inductive Logic Programming approach called Logical Vision (LV) was shown to overcome some of these limitations. LV uses Meta-Interpretive Learning combined with low-level extraction of high-contrast points sampled from the image to learn recursive logic programs describing the image. In published work LV was demonstrated capable of high-accuracy prediction of classes such as *regular polygon* from a small number examples of images where the compared statistical learning algorithms gave near random prediction given hundreds of instances. LV has so far only been applied to noise-free, artificially generated images. This paper extends LV by using a) richer background knowledge such as light reflection that can itself be learned and used for resolving visual ambiguities, which cannot be easily modeled using statistical approaches, b) a wider class of background models representing classical 2D shapes such as circles and ellipses, c) primitive-level statistical estimators to handle noise in real images, Our results indicate that in real images the new noise-robust version of LV using a single example (ie one-shot LV) converges to an accuracy at least comparable to thirty-shot statistical machine learner on the prediction of hidden light sources. Moreover, we demonstrate that the learned theory can be used to identify ambiguities in the convexity/concavity of objects such as craters.

1 Introduction

Galileo's *Siderius Nuncius* [11] describes the first ever telescopic observations of the moon. Using sketches of shadow patterns Galileo conjectured the existence of mountains containing hollow areas (i.e. craters) on a celestial body previously thought perfectly spherical. His reasoned description, derived from a handful of observations, relies on a knowledge of i) classical geometry, ii) straight line movement of light and iii) the Sun as a light source. This paper investigates use of Inductive Logic Programming (ILP) [27] to derive such hypotheses from a small set of real images. Figure 1 illustrates part of the generic background knowledge used by ILP for interpreting object convexity.



a) b) c) d)

e) $light(X, X).$
$light(X, Y) :- reflect(X, Z), light(Z, Y).$

Fig. 1: Interpretation of light source direction: a) Waxing crescent moon (Credit: UC Berkeley), b) Concave/Convex illusion caused by the viewer's assumption about the light source location, c) Concave and d) Convex photon reflection models, e) Prolog recursive model of photon reflection

Figure 1a shows an image of the crescent moon in the night sky, in which convexity of the overall surface implies the position of the Sun as a hidden light source beyond the lower right corner of the image. Figure 1b shows an illusion caused by the viewer's assumption about where the light source is. Assuming the light source is above makes the top right and bottom left circles appear *convex* and the other circles *concave*. Assuming the light source is below makes the top left and bottom right circles appear *convex* and the other circles *concave*. Figure 1c shows how interpretation of a *convex* feature, such as a mountain, comes from illumination of the *right* side of a convex object. Figure 1d shows that perception of a *concave* feature, such as a crater, comes from illumination of the *left* side. Figure 1e shows how Prolog background knowledge encodes a recursive definition of the reflected path of a photon.

This paper explores the phenomenon of knowledge-based perception using an extension of Logical Vision (LV) [7] based on Meta-Interpretive Learning (MIL) [26, 6]. In the previous work LV was shown to accurately learn a variety of polygon classes from artificial images with low sample requirements compared to statistical learners. In this paper we propose a noise-robust version of LV provided with basic generic background knowledge about radiation and reflection of photons to inform the generation of hypotheses in the form of logic programs based on evidence sampled from a single real image. Our experiments show that LV converges to high accuracy in prediction of light source position. This compares with Support Vector Machines (SVMs) using stronger supervision only achieving similar accuracy at least after more than 30 images. Clearly such image sample requirements compare poorly with both LV and human vision for situations in which effective learning must be achieved from small samples.

The main contributions of this paper are extending LV [7] by using 1) richer background knowledge in the form of a simple but generic recursive theory of light reflection and also demonstrating that this recursive theory of light can itself be learned using MIL in order to resolve visual ambiguities which cannot be easily modelled using statistical approaches, 2) a wider class of background models representing classical 2D shapes such as circles and ellipses, 3) primitive-level statistical estimators to handle noise in real images and demonstrating that the extended LV using a single example (i.e. one-shot LV) converges to an accuracy at least comparable to thirty-shot statistical learner.

The paper is organised as follows. Section 2 describes related work. The theoretical framework for LV is provided in Section 3. Section 4 describes the implementation of LV, including the recursive background knowledge for describing radiation and reflection of light. Experiments on predicting the light source direction in images of the moon and microscopic images of illuminated micro-organisms are described in Section 5. In Section 6 we discuss how the approach interprets convexity, concavity and visual illusions. Finally, we conclude and discuss further work in Section 7.

2 Related work

Statistical machine learning based on low-level feature extraction has been increasingly successful in image classification [30]. However, high-level vision, involving interpretation of objects and their relations in the external world, is still relatively poorly understood [4]. Since the 1990s *perception-by-induction* [13] has been the dominant model within computer vision, where human perception is viewed as inductive inference of hypotheses from sensory data. The idea originated in the work of the 19th century physiologist Hermann von Helmholtz [15]. The approach described in this paper is in line with *perception-by-induction* in using ILP for generating high-level perceptual hypotheses by combining sensory data with a strong bias in the form of explicitly encoded background knowledge. Whilst Gregory [12] was one of the earliest to demonstrate the power of the Helmholtz's perception model for explaining human visual illusion, recent experiments [14] show Deep Neural Networks fail to reproduce human-like perception of illusion. This contrasts with results in Section 6, in which LV achieves analogous outcomes to human vision.

Shape-from-shading [16, 34] is a key computer vision technology for estimating low-level surface orientation in images. Unlike our approach for identifying concavities and convexities, shape-from-shading generally requires observation of the same object under multiple lighting conditions. By using background knowledge as a bias we reduce the number of images for accurate perception of high-level shape properties such as the identification of convex and concave image areas.

ILP has previously been used for learning concepts from images. For instance, in [3, 1] object recognition is carried out using existing low-level computer vision approaches, with ILP being used for learning general relational concepts from this already symbolised starting point. By contrast, LV [7] uses ILP and abductive perception technique [31] to provide a bridge from very low-level primitives, such as high contrast points, to higher-level interpretation of objects such as shapes. ILP also has been used for 3D scene analysis [10, 25] with 3D point cloud data, however there was no comparison made to statistical learning and image ambiguity not addressed.

The present paper extends the earlier work on LV by implementing a noise-proofing technique, applicable to real images, and extending the use of background knowledge radiation to allow the identification of objects such as light sources, not directly identifiable within the image itself. Moreover, this work shows that by considering generic knowledge about radiation, LV can invent generic high-level concepts applicable to many different images including concavity, convexity and light reflection, enabling 2D image analysis to learn a 3D concept with ambiguity handled.

One-shot learning of concepts from images using probabilistic program induction is discussed in [18, 19]. However, unlike the approach in this paper, the images are relatively simple and artificially generated and learning involves parameter estimation for a given program schema, rather than a search through general program space, relative to incrementally generated background knowledge.

Various statistics-based techniques making use of high-level vision have been proposed for one- or even zero-shot learning [29, 32]. They usually start from an existing model pre-trained on a large corpus of instances, and then adapt the model to data with unseen concepts. Approaches can be separated into two categories. The first exploits a mapping from images to a set of semantic attributes, then high-level models are learned based on these attributes [20, 23, 29]. The second approach uses statistics-based methods, pre-trained on a large corpus, to find localised attributes belonging to objects but not the entire image, and then exploits the semantic or spatial relationships between the attributes for scene understanding [17, 21, 9]. Unlike these approaches, we focus on one-shot from scratch, i.e. high-level vision based on just *very low-level primitives* such as high contrast points.

3 Framework

The framework for LV is a special case of MIL.

3.1 Meta-Interpretive Learning

Given background knowledge B and examples E the aim of a MIL system is to learn a hypothesis H such that $B, H \models E$, where $B = B_p \cup M$, B_p is a set of Prolog definitions and M is a set of *metarules* (see Figure 2). MIL [25, 26, 5, 24, 6] is a form of ILP based on an adapted Prolog meta-interpreter. A standard Prolog meta-interpreter proves goals by repeatedly fetching first-order clauses whose heads unify with the goal. By contrast, a MIL learner proves a set of examples by fetching higher-order metarules (Figure 2) whose heads unify with the goal. The resulting meta-substitutions are saved, allowing them to be used to generate a hypothesised program which proves a the examples by substituting the meta-substitutions into corresponding metarules.

Name	**Metarule**
PropObj1	$P(obj1) \leftarrow$
PropObj2	$P(obj2) \leftarrow$
PropLight	$P(light) \leftarrow$
Conjunct3	$P(x,y,z) \leftarrow Q(x,y,z), R(x,y,z)$
Chain3	$P(u,x,y) \leftarrow Q(u,x,z), R(u,z,y)$
Chain32	$P(u,x,y) \leftarrow Q(u,x,z), R(z,y)$
PrePost3	$P(x,y,z) \leftarrow Q(x,y), R(x), S(z)$

Fig. 2: Metarules used in this paper. Uppercase letters P, Q, R, S denote existentially quantified variables. Lowercase letters u, x, y, and z are universally quantified.

MIL sample complexity Use of metarules and background knowledge helps minimise the number of clauses n of the minimal consistent hypothesis H and consequently the number of examples m required to achieve error below ϵ bound. As shown in [6], n dominates the lower bound for m[4].

[4] $m \geq \frac{n\ ln|M| + p\ ln(3n) + ln\frac{1}{\delta}}{\epsilon}$ for p predicates and M metarules

3.2 Logical Vision

In LV [7], the background knowledge B, in addition to Prolog definitions, contains a set of one or more named images I. The examples describe properties associated with I.

4 Implementation

Our implementation of LV, called $LogVis$, is shown in Algorithm 1. Note that $LogVis$ calls the MIL system $Metagol_{AI}$ [6] for inductive inference over given images and background knowledge.

4.1 Meta-interpretion in real images

The interpretion of image is based on the objectDetection procedure in Algorithm 1, which is an extension of that in LV for polygon learning [7]. The new version supports noise-robustness in real image analysis. The basic primitive of LV is *edge_point([X,Y])*, which decides if an image pixel belongs to the edge of a foreground object. An example of object detection is shown in Figure 3.

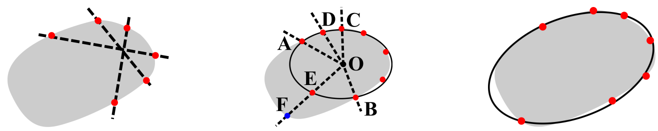
The results of *edge_point/1* are affected by noise in real images. We address this by using primitive predicates which call statistical models to perform evaluation on images. For example, *edge_point/1* calls a pre-trained statistical model which classifies pixels into background and foreground using Gaussian models or image segmentation.

Objects of interest in microscopic and telescopic images are often composed of curves. Therefore, bounding boxes, popular for object representation in computer vision, include surrounding areas that do not belong to the objects, resulting in reduced accuracy. Consequently we use ellipse and circle models estimated from sets of edge points (see Figure 3).

Algorithm 1: $LogVis(I, B)$

Input : Training images I; Background knowledge B.
Output : Hypothesised logic program H.

Candidates = Φ;
for each *labelled image* $i \in I$ **do**
    Angles = Φ;
    /* Object & highlight detection */
    for $t \in [1, T]$ **do**
        Obj = $objectDetection(i)$;
        α = argmax$_{Angle}$ $contrast$($split$(Obj, Angle));
        Angles = $append$(Angles, α);
    end
    /* Highlight angle */
    HAngle = $mode$(Angles);
    /* Light source angle */
    LAngle = $label(i)$;
    /* Call $Metagol_{AI}$ to learn a model */
    Model = $Metagol_{AI}(B$, HAngle, LAngle);
    Candidates = add(Model,Candidates);
end
Return(H = $best$(Candidates));



a) b) c)

Fig. 3: Object detection: a) Sampled lines with edge points; b) Fitting of initial ellipse centred at O. Hypothesis tested using new edge points halfway between existing adjacent points. c) Revised hypothesis tested until hypothesis passes test.

Detected objects take the form *elps(Centre, Parameter)* or *circle(Centre, Radius)* where $Centre = [X, Y]$ is the object's centre, $Parameter = [A, B, Tilt]$ are the axis lengths and tilting angle and $Radius$ is the circle radius.

To estimate light source direction LogVis (line 6) cuts the object in half at different angles, and returns the angle α which maximises brightness contrast between the halves, where $\alpha \in \{1..12\}$ is a clock face angle. Since noise may cause object detection to fail, LV repeats the process T times and returns $HAngle$ as the mode of $\{\alpha\}$, (line 4 to 9). The output H is the clock angle.

Background knowledge for $Metagol_{AI}$ is shown in Figure 4. Together with the metarules in Figure 2, $Metagol_{AI}$ can learn an abductive theory (line 11 in Algorithm 1).

Primitives	Compiled BK
prim(light_source_angle/3). *% supervision*	highlight(obj1,obj2). *% $obj2$ is highlight on $obj1$*
prim(highlight/2). *% abduced by LV*	opposite_angle(3, 9). opposite_angle(9, 3).
prim(opposite_angle/2).	opposite_angle(12, 6). opposite_angle(6, 12).

Fig. 4: Background knowledge for $Metagol_{AI}$

When a dataset has more than one example, LV runs the entire one-shot learning process for a random example, and returns the most accurate hypothesis on the rest of training set (line 14).

5 Experiments

This section describes experiments comparing one-shot LV with multi-shot statistics-based learning[5] on real image datasets. Here we investigate the following null hypothesis:

Null hypothesis: One-shot LV cannot learn models with accuracy comparable to thirty-shot statistics-based learning on real images.

5.1 Materials

We collected two real image datasets for the experiments: 1) **Protists** drawn from a (coloured) microscope video of a *Protist* micro-organism, and 2) **Moons** a collection of (grey-scaled) images of the moon drawn from Google images. To formulate a classification problem, we use 12 clock angles[6] to descretisize the learning target, light source angle. The datasets consist of 30 images for each angle, providing a total of 360 images. Each image contains one of four labels as follows: $North = \{11, 12, 1\}$ clocks, $East = \{2, 3, 4\}$ clocks, $South = \{5, 6, 7\}$ clocks, and $West = \{8, 9, 10\}$ clocks, as shown in Fig 5.

[5] Data and code at https://github.com/haldai/LogicalVision2

[6] Clock face angle between 12 and each hour position in $\{1..12\}$.

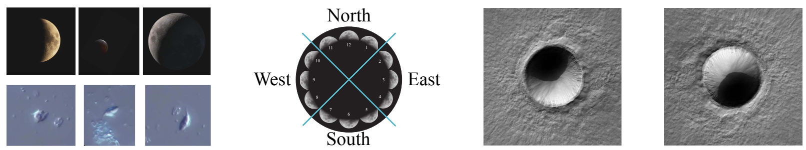


Fig. 5: Illustrations of data: a) Examples of the datasets, b) Four classes for twelve light source positions, c) Crater on Mars (Credit: NASA/JPL/University of Arizona), d) 180° rotated crater.

5.2 Methods

The aim is to learn a model to predict the correct category of light source angle from real images. For each dataset, we randomly divided the 360 images into training and test sets, with 128 and 232 examples respectively. To evaluate the performance, the models were trained by randomly sampling 1, 2, 4, 8, 16, 32, 64 and 128 images from the training set. The sequences of training and test instances are shared by all compared methods. The random partition of data and learning are repeated 5 times.

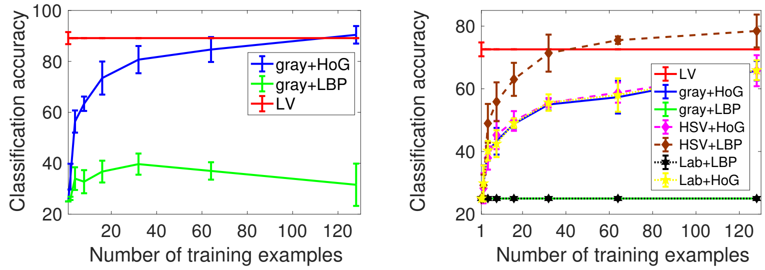
Logical Vision In the experiments, we used the grey intensity of both image datasets for LV. The hyper-parameter T in Algorithm 1 is set at 11 by validating one-shot learned models on the rest of the training data. To handle image noise, we use a statistics-based estimator for predicate *edge_point/1*. When *edge_point([X,Y])* is called, a vector of colour distribution (histogram of grey-scale value) of the 10×10 region centred at (X,Y) is calculated, then the statistical model is applied to determine whether this vector represents an edge point. The statistical model is trained from 5 randomly sampled images in the training set by providing the bounding box of the objects.

Statistics-based Classification The experiments with statistics-based classification were conducted in different colour spaces combined with various features. Firstly, we performed feature extraction to transform images into fixed length vectors. Next SVMs (libSVM [2]) with non-linear kernel were applied to learn a multiclass-classifier model. Parameters of the SVM are chosen by cross-validation on the training set.

Like LV, we used grey intensity from both image datasets for the experiments. For the coloured *Protists* dataset, we transformed the images to **HSV** and **Lab** colour spaces to improve the performance.

Since the image sizes in the dataset are irregular, during the object detection stage of LV, we used computer graphic techniques (e.g. curve fitting) to extract the main objects and unified them into same sized patches for feature extraction. For the feature extraction process, we avoided descriptors which are insensitive to scale and rotation, instead choosing the luminance-sensitive features below.

- **HOG**: The Histogram of Oriented Gradient (HOG) [8] is known as its capability of describing the local gradient orientation in an image, and widely used in computer vision and image processing for the purpose of object detection.
- **LBP**: Local binary pattern (LBP) [28] is a powerful feature for texture classification by converting the local texture of an image into a binary number.



a) *Moons* b) *Protists*

Fig. 6: Classification accuracy on the two datasets.

Remark Despite our best efforts it proved impossible to make testing entirely fair. In the *Moons* task, LV and the compared statistics-based approach both used geometrical background knowledge for fitting circles (though in different forms) during object extraction. However, in the *Protists* task, the noise in images always caused poor performance in automatic object extraction for the statistics-based method. Therefore, *we provided additional supervision to the statistics-based method consisting of bounding boxes for the main objects in both training and test images during feature extraction*. By comparison LV discovers the objects without any supervision.

5.3 Results

Figure 6a shows the results for *Moons*. Note that performance of the statistics-based approach only surpasses one-shot LV after 100 training examples. In this task, background knowledge involving circle fitting exploited by LV and statistics-based approaches are similar, though low-level feature used by statistics-based approach are first-order information (grey-scale gradients), which is stronger than the zero-order information (grey-scale value) used by LV.

Results on *Protists* are shown in Figure 6b. After 30+ training examples only one statistics-based approach outperforms one-shot LV. Since the statistics-based approaches have additional supervision (bounding box of main object) in the experiments, improved performance is unsurprising.

The results of LV in Figure 6 form a horizontal lines. When the number of training examples exceeds one, LV performs multiple one-shot learning and selects the best output, which we found is always in the same equivalent class in LV's hypothesis space. This suggests LV learns the optimal model in its hypothesis space from a single example, while the mis-classification are resulted by the noise in LV's object-detection stage. The learned program is shown in Figure 7a.

By comparison the statistics-based approaches require 40 or even 100 more training examples to reach similar accuracy, which refutes the null hypothesis.

LV is implemented in SWI-Prolog [33] with multi-thread processing. Experiments were executed on a laptop with Intel i5-3210M CPU (2.50GHz), the time costs of object discovery are 9.5 seconds and 6.4 seconds per image on *Protists* and *Moons* dataset respectively; the average running time of $Metagol_{AI}$ procedure is 0.001 second on both datasets.

```
clock_angle(A,B,C):-
    clock_angle1(A,B,D),
    light_source_angle(A,D,C).
clock_angle1(A,B,C):-
    highlight(A,B),
    clock_angle2(A),clock_angle3(C).
clock_angle2(obj1).
clock_angle3(light).
```

a)

```
clock_angle(A,B,C):-
    clock_angle1(A,B,D),
    clock_angle4(A,D,C).
clock_angle4(A,B,C):-
    light_source_angle(A,B,D),
    opposite_angle(D,C).
```

b)

Fig. 7: Program learned by LV: a) with background knowledge about lighting, we can understand that the invented predicate *clock_angle2* stands for *convex*, *clock_angle3* stands for *light_source_name*. b) Learned program when concave objects are given as training examples, where clock_angle1 is same with a).

Protists and *Moons* contain only convex objects. If instead we provide images with concave objects (such as Figure 5c and d), LV learns a program such as Figure 7b. Here the invented predicate *clock_angle2/1* can be interpreted as *concave*.

6 Discussion: Learning ambiguity

Figure 5c and 5d shows two images of a crater on Mars, where Figure 5d is a $180°$ rotated image of Figure 5c. Human perception often confuses the convexity of the crater in such images[7]. This phenomenon, called the *crater/mountain illusion*, occurs because human vision usually interprets pictures under the default assumption that the light is from the top of the image.

LV can use MIL to perform abductive learning. We show below that incorporation of generic recursive background knowledge concerning light enables LV to generate multiple mutually inconsistent perceptual hypotheses from real images. To the authors' knowledge, such ambiguous prediction has not been demonstrated previously with machine learning.

Recall the learned programs from Figure 7 from the previous experiments. If we rename the invented predicates we get the general theory about lighting and convexity shown in Figure 8.

```
clock_angle(O,H,A):-
    highlight(O,H),convex(O),light_source(L),
    light_source_angle(O,L,A).
clock_angle(O,H,A):-
    highlight(O,H),concave(O),light_source(L),
    light_angle(O,L,A1),opposite(A1,A).
```

Fig. 8: Interpreted BK learned by LV.

Now we can use the program as a part of interpreted background knowledge for LV to do abductive learning, where the abducible predicates and the rest of background knowledge are shown in Figure 9.

[7] http://www.universetoday.com/118616/do-you-see-a-mountain-or-a-crater-in-this-picture/

Abducibles	Interpreted BK
prim(convex/1). prim(concave/1). prim(light_source/1). prim(light_angle/3).	highlight(X,Y):- contains(X,Y),brighter(Y,X),light_source(L), light_path(L,R),reflector(R),light_path(R,O), observer(O).
Compiled BK	
% "obj1" is an object abduced from image; "obj2" is the brighter part of "obj1"; *% "observer" is the camera* contains(obj1,obj2). brighter(obj2,obj1). observer(observer). reflector(obj2). light_path(X,X). light_path(X,Y):-unobstructed(X,Z), light_path(Z,Y).	

Fig. 9: Background knowledge for learning ambiguity from images.

If we input Figure 5c to LV, it will output four different abductive hypotheses for the image, as shown in Figure 10[8]. From the first two results we see that, by considering different possibilities of light source direction, LV can predict that the main object (which is the crater) is either convex or concave, which shows the power of learning ambiguity. The last two results are even more interesting: they suggest that *obj2* (the highlighted part of the crater) might be the light source as well, which indeed is possible, though seems unlikely.[9]

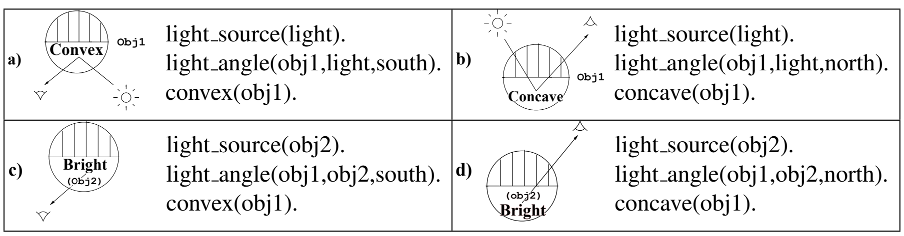


Fig. 10: Depiction and output hypotheses abduced from Figure 5c.

7 Conclusions and further work

Human beings learn visual concepts from single image presentations (so-called one-shot-learning) [18]. This phenomenon is hard to explain from a standard Machine Learning perspective, given that it is unclear how to estimate any statistical parameter from a single randomly selected instance drawn from an unknown distribution. In this paper we show that learnable generic logical background knowledge can be used to generate high-accuracy logical hypotheses from single examples. This compares with similar demonstrations concerning one-shot MIL on string transformations [22] as well as previous concept learning in artificial images [7]. The experiments in Section 5 show that the LV system can accurately identify the position of a light source from a single real image, in a way analogous to scientists such as Galileo, observing the moon for the first time through a telescope or Hook observing micro-organisms for the first time

[8] Code also at https://github.com/haldai/LogicalVision2

[9] The result can be reproduced and visualised by the example in Logical Vision 2 repository.

through a microscope. In Section 6 we show that logical theories learned by LV from labelled images can also be used to predict concavity and convexity predicated on the assumed position of a light source.

As future work, we aim to investigate broader sets of visual phenomena which can naturally be treated using background knowledge. For instance, the effects of object obscuration; the interpretation of shadows in an image to infer the existence of out-of-frame objects; the existence of unseen objects reflected in a mirror found within the image. All these phenomena could possibly be considered in a general way from the point of view of a logical theory describing reflection and absorption of light. We will also investigate the use of universal meta-rules similar to those used in [5]. Future work also includes the use of probabilistic representation.

The authors believe that LV has long-term potential as an AI technology with the potential for unifying the disparate areas of logical based learning with visual perception.

References

1. Antanas, L., van Otterlo, M., Oramas Mogrovejo, J., Tuytelaars, T., De Raedt, L.: There are plenty of places like home: Using relational representations in hierarchies for distance-based image understanding. Neurocomputing 123, 75–85 (2014)
2. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011)
3. Cohn, A., Hogg, D., Bennett, B., Galata, A., Magee, D., Santos, P.: Cognitive vision: Integrating symbolic qualitative representations with computer vision. In: Cognitive Vision Systems, pp. 221–246. Springer, Berlin (2006)
4. Cox, D.: Do we understand high-level vision? Current opinion in neurobiology 25, 187–193 (2014)
5. Cropper, A., Muggleton, S.: Logical minimisation of meta-rules within meta-interpretive learning. In: Proceedings of the 24th International Conference on Inductive Logic Programming. pp. 65–78. Springer-Verlag (2015)
6. Cropper, A., Muggleton, S.: Learning higher-order logic programs through abstraction and invention. In: Proceedings of the 25th International Joint Conference Artificial Intelligence. pp. 1418–1424 (2016)
7. Dai, W.-Z., Muggleton, S., Zhou, Z.-H.: Logical Vision: Meta-interpretive learning for simple geometrical concepts. In: Late Breaking Paper Proceedings of the 25th International Conference on Inductive Logic Programming. pp. 1–16. CEUR (2015)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the 13rd IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 886–893. IEEE Computer Society, San Diego, CA (2005)
9. Duan, K., Parikh, D., Crandall, D.J., Grauman, K.: Discovering localized attributes for fine-grained recognition. In: Proceedings of the 25th IEEE Conference on Computer Vision and Pattern Recognition. pp. 3474–3481. IEEE Computer Society, Providence, RI (2012)
10. Farid, R., Sammut, C.: Plane-based object categorisation using relational learning. Machine Learning 94(1), 3–23 (2014)
11. Galilei, G.: The Herald of the Stars (1610), english translation by Edward Stafford Carlos, Rivingtons, London, 1880; edited by Peter Barker, Byzantium Press, 2004
12. Gregory, R.: Concepts and Mechanics of Perception. Duckworth, London (1974)
13. Gregory, R.: Eye and Brain: The Psychology of Seeing. Oxford University Press, Oxford (1998)

14. Heath, D., Ventura, D.: Before a computer can draw, it must first learn to see. In: Proceedings of the 7th International Conference on Computational Creativity. pp. 172–179 (2016)
15. von Helmholtz, H.: Treatise on Physiological Optics Volume 3. Dover Publications, New York (1962), originally published in German in 1825
16. Horn, B.: Obtaining shape from shading information. MIT Press (1989)
17. Hu, R., Xu, H., Rohrbach, M., Feng, J., Saenko, K., Darrell, T.: Natural language object retrieval. In: Proceedins of the 29th IEEE Conference on Computer Vision and Pattern Recognition. pp. 4555–4564. IEEE Computer Society, Las Vegas, NV (2016)
18. Lake, B., Salakhutdinov, R., Gross, J., Tenenbaum, J.: One shot learning of simple visual concepts. In: Proceedings of the 33rd Annual Conference of the Cognitive Science Society. pp. 2568–2573 (2011)
19. Lake, B., Salakhutdinov, R., Tenenbaum, J.: Human-level concept learning through probabilistic program induction. Science 350, 1332–1338 (2015)
20. Lampert, C.H., Nickisch, H., Harmeling, S.: Attribute-based classification for zero-shot visual object categorization. IEEE Transactions on Pattern Analysis and Machine Intelligence 36(3), 453–465 (2014)
21. Li, Z., Gavves, E., Mensink, T., Snoek, C.G.M.: Attributes make sense on segmented objects. In: Proceedings of 13th European Conference on Computer Vision Part IV. pp. 350–365. Springer, Zurich, Switzerland (2014)
22. Lin, D., Dechter, E., Ellis, K., Tenenbaum, J., Muggleton, S.: Bias reformulation for one-shot function induction. In: Proceedings of the 23rd European Conference on Artificial Intelligence (ECAI 2014). pp. 525–530. IOS Press, Amsterdam (2014)
23. Mensink, T., Verbeek, J.J., Csurka, G.: Learning structured prediction models for interactive image labeling. In: The 24th IEEE Conference on Computer Vision and Pattern Recognition. pp. 833–840. IEEE Computer Society, Colorado Springs, CO (2011)
24. Muggleton, S., Lin, D., Chen, J., Tamaddoni-Nezhad, A.: Metabayes: Bayesian meta-interpretative learning using higher-order stochastic refinement. In: Zaverucha, G., Costa, V.S., Paes, A.M. (eds.) Proceedings of the 23rd International Conference on Inductive Logic Programming. pp. 1–17. Springer-Verlag, Berlin (2014)
25. Muggleton, S., Lin, D., Pahlavi, N., Tamaddoni-Nezhad, A.: Meta-interpretive learning: application to grammatical inference. Machine Learning 94, 25–49 (2014)
26. Muggleton, S., Lin, D., Tamaddoni-Nezhad, A.: Meta-interpretive learning of higher-order dyadic datalog: Predicate invention revisited. Machine Learning 100(1), 49–73 (2015)
27. Muggleton, S., Raedt, L.D., Poole, D., Bratko, I., Flach, P., Inoue, K.: ILP turns 20: biography and future challenges. Machine Learning 86(1), 3–23 (2011)
28. Ojala, T., Pietikainen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
29. Palatucci, M., Pomerleau, D., Hinton, G., Mitchell, T.M.: Zero-shot learning with semantic output codes. In: Advances in Neural Information Processing Systems 22, pp. 1410–1418. Curran Associates Inc. (2009)
30. Poppe, R.: A survey on vision-based human action recognition. Image and vision computing 28(6), 976–990 (2010)
31. Shanahan, M.: Perception as abduction: Turning sensor data into meaningful representation. Cognitive Science 29(1), 103–134 (2005)
32. Vinyals, O., Blundell, C., Lillicrap, T.P., Kavukcuoglu, K., Wierstra, D.: Matching networks for one shot learning. CoRR abs/1606.04080 (2016)
33. Wielemaker, J., Schrijvers, T., Triska, M., Lager, T.: SWI-Prolog. Theory and Practice of Logic Programming 12(1-2), 67–96 (2012)
34. Zhang, R., Tai, P., Cryer, J., Shah, M.: Shape-from-shading: a survey. IEEE transactions on pattern analysis and machine intelligence 21(8), 670–706 (1999)