Advances in Cognitive Systems 6 (2018) 65–83 Submitted 6/2018; published 9/2018

Semantic Analysis of Reflectional Visual Symmetry: A Human-Centered Computational Model for Declarative Explainability

Jakob Suchan JSUCHAN@UNI-BREMEN.DE
Mehul Bhatt MEHUL.BHATT@ORU.SE
Srikrishna Varadarajan KRISH@CODESIGN-LAB.ORG
CoDesign Lab EU, Örebro University, Sweden / Cognitive Vision, University of Bremen, Germany

Seyed Ali Amirshahi S.ALI.AMIRSHAHI@NTNU.NO
The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology, Norway

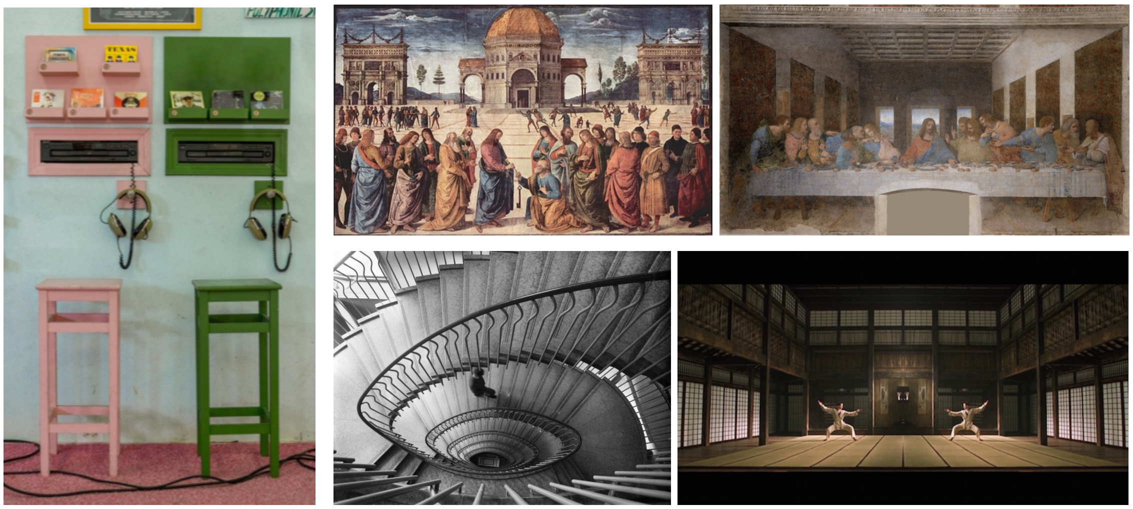
Stella Yu STELLAYU@BERKELEY.EDU
ICSI Vision Group, International Computer Science Institute, Berkeley, CA 94704 USA

Abstract

We present a computational model for the semantic interpretation of symmetry in naturalistic scenes. Key features include a human-centered representation, and a declarative, explainable interpretation model supporting deep semantic question-answering founded on an integration of methods in knowledge representation and deep learning based computer vision. In the backdrop of the visual arts, we showcase the framework's capability to generate human-centered, queryable, relational structures, also evaluating the framework with an empirical study on the human perception of visual symmetry. Our framework is driven by the application and integration of methods for foundational vision, knowledge representation, and reasoning to the arts, while incorporating evidence from the psychological and social sciences.

1. Introduction

Visual symmetry as an aesthetic and stylistic device has been employed by artists across a spectrum of creative endeavours concerned with visual imagery in some form, such as painting, photography, architecture, film and media design. Symmetry in visual art and beyond is often linked with elegance, beauty, and is associated with attributes such as being well proportioned and well balanced (Weyl, 1952). Closer to the "visual imagery" and "aesthetics" centered scope of this paper, symmetry has been employed by visual artists going back to the masters Giorgione, Titian, Raphael, da Vinci, and continuing into modernity with Dali and other contemporary artists, as shown in Figure 1.



(a) (b)

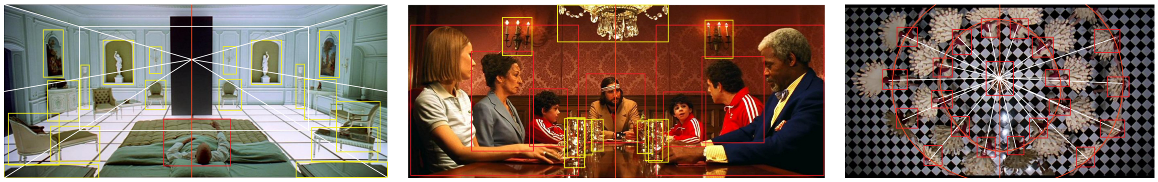
Figure 1. The perception of symmetry. (a) Symmetry perception influenced by visual features, conceptual categories, semantic layering, and nuances of individual differences in perception, and (b) examples for symmetry in visual arts: *"Delivery of the Keys"* (ca.1481) by Perugino, *"The Last Supper"* (1495-98) by Leonardo Da Vinci, *"View of the grand staircase at La Rinascente in Rome, designed by Franco Albini and Franca Helg"* (1962) by Giorgio Casali, and *"The Matrix"* (1999) by the Wachowski Brothers.

Visual Symmetry: Perception and Semantic Interpretation. There exist at least four closely related points of view pertaining to symmetry, namely, the physical, mathematical, pyschological, and aesthetical points of view. As Molnar and Molnar (1986) articulate:

> But perceptual symmetry is not always identical to the symmetry defined by the mathematicians. A symmetrical picture is not necessarily symmetrical in the mathematical sense… Since the aesthetical point of view is strictly linked to the perceptive system, in examining the problems of aesthetics we find ourselves dealing with two distinct groups of problems: (1) the problem of the perception of symmetry; (2) the aesthetical effect of the perception of a symmetrical pattern.

Indeed, the high-level semantic interpretation of symmetry in naturalistic visual stimuli by humans is a multi-layered perceptual phenomenon operating at several interconnected cognitive levels involving spatial organization, visual features, semantics, and individual differences (Section 2.1; and Figure 1a). Consider the select examples from movie scenes in Figure 2:

- In the shot from "2001: A Space Odyssey" (Figure 2a) a center-perspective is being applied for staging the scene. The symmetry here is obtained by this, as well as by the layout of the room, the placement of the furniture, and the decoration of the room. In particular, the black obelisk in the center of the frame emphasizes the center-perspective regularly used by Kubrick, with the bed (and person) being positioned directly on the central axis.



(a) (b) (c)

Figure 2. Symmetrical structure in visual arts illustrated by select scenes from films: (a) *"2001: A Space Odyssey"* (1968) by Stanley Kubrick, (b) *"The Royal Tenenbaums"* (2001) by Wes Anderson, and (c) *"The Big Lebowski"* (1998) by Joel and Ethan Coen.

- Wes Anderson is staging his shot from "The Royal Tenenbaums" (Figure 2b) around a central point, but, unlike Kubrick's shot, focuses on the people involved in it. Even though the visual appearance of the characters differs considerably, the spatial arrangement and the semantic similarity of the objects in the shot creates symmetry. Furthermore, the gazing direction of the characters, i.e., people on the right facing left and people on the left facing right, adds to the symmetrical appearance.
- In "The Big Lebowski" (Figure 2c), Joel and Ethan Coen use symmetry to highlight the surreal character of a dream sequence; the shot in Figure 2c uses radial symmetry composed of a group of dancers, shot from above, moving around the center of the frame in a circular motion. This is characterized by moving entities along a circular path and center point, and the perceptual similarity of the dancers' appearance.

The development of computational cognitive models focusing on a human-centered – *semantic, explainable* – interpretation of visuo-spatial symmetry presents a formidable research challenge demanding an interdisciplinary (mixed-methods) approach at the interface of cognitive science, vision and AI, and visual perception focused human-behavioral research. Broadly, our research is driven by addressing this interdisciplinarity, with an emphasis on developing integrated reasoning and visual processing for applications such as automated archive annotation, preprocessing for qualitative analysis, and studies in visual perception.

Key Contributions. The core focus of the paper is to present a computational model with the capability to generate semantic, explainable interpretation models for the analysis of visuo-spatial symmetry. The explainability is founded on a domain-independent, qualitative-quantitive representation of visuo-spatial relations that characterizes symmetry in declarative terms. We also report on a qualitative evaluation with humans, whereby subjects rank their subjective perception of visual symmetry for a set stimuli using (qualitative) distinctions. The broader implications are twofold: (1) the paper demonstrates the integration of vision and semantics by combining knowledge representation and reasoning with low-level (deep learning based) visual processing; and (2) from an applied viewpoint, the developed methodology can serve as the technical backbone for assistive and analytical technologies for visual media studies, from the viewpoint of psychology, aesthetics, and cultural heritage.

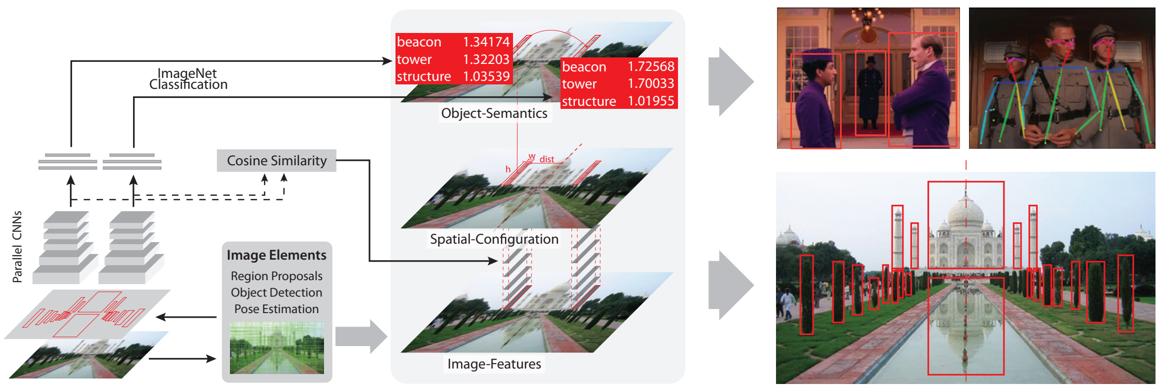


Figure 3. A computational model of multi-level semantic symmetry.

2. The Semantics of Symmetry

Symmetry in visual imagery denotes that an image is invariant to certain types of transformation. For instance, reflectional symmetry occurs when the image does not change if it is mirrored along a specific axis of symmetry. Other types of symmetry include rotational symmetry and translational symmetry. Perfect symmetry can be easily detected based on image level features by comparing pixels in the image; however, in natural images (e.g., coming from the visual arts), perfect symmetry is very rare and approximations are used as a stylistic device, with it occurring only in some parts of the image. To address this, we focus on developing a semantic model capable of interpreting symmetrical structures in such images.

2.1 A Multi-Level Semantic Characterization

From the viewpoint of perceptual and aesthetic considerations, key aspects for interpreting visual-spatial symmetry (in the scope of this paper) include:

(S1) **Spatial organization**: High-level conceptual categories identifiable from geometric constructions by way of arbitrary shapes, relative orientation and placement, size of geometric entities, relative distance, and depth;

(S2) **Visual features**: Low-level visual features and artifacts emanating directly from color, texture, light, and shadow;

(S3) **Semantic layers**: Semantic layering and grouping based on natural scene characteristics involving, for instance, foreground and background, conceptual similarity, relative distance, and perceived depth, and commonsense knowledge not directly available in the stimulus;

(S4) **Individual differences**: Grounding of the visual features in the socio-cultural semiotic landscape of the perceiver (i.e., contextual and individualized nuances in perception and sense making).

We develop a multi-level characterization of symmetry aimed at analyzing (reflectional) symmetry. In this paper, visual symmetry encompasses three layers:

L1. *Symmetrical (spatial) composition*: Spatial arrangement of objects in the scene with respect to a structural representation of position, size, orientation, and the like;

L2. *Perceptual similarity*: Perceptual similarity of features in symmetrical image patches based on the low-level feature based appearance of objects, such as colour, shape, and patterns;

L3. *Semantic similarity*: Similarity of semantic categories of the objects in symmetrical image patches, such as people, object types, and properties of these objects, including people's gazing direction and foreground / background relations.

Our characterization serves as the foundation for analyzing and interpreting symmetrical structures in images; in particular, it can be used to identify not only elements of the image that support the symmetrical structure, but also those parts of the image that break the symmetry. This may be used for investigating the use of balance and imbalance in the visual arts and for analyzing how this can be used to guide viewers' attention in the context of visual saliency.

2.2 A Model of Reflectional Symmetry

Figure 3 depicts the computational model presented in this paper, which focuses on reflectional symmetry in the composition of the image based on layers L1–L3. That is, we will investigate image properties based on spatial configuration, low-level feature similarity, and semantic similarity. To this end, we extract three types of image elements $\mathcal{E}_{1\cup 2\cup 3} = \{e_0, \ldots, e_n\}$ from the image:

($\mathcal{E}_1$) *Image patches* are extracted using selective search, as described by Uijlings et al. (2013), resulting in structural parts of the image, potential objects, and object parts;

($\mathcal{E}_2$) *People and objects* are detected in the image using YOLO object detection (Redmon et al., 2016);

($\mathcal{E}_3$) *Human body poses*, consisting of body joints and facing direction, are extracted using methods for human pose estimation (Cao et al., 2017).

Potential symmetrical structures in the image are defined over the image elements $\mathcal{E}$ in terms of identified pairs of image elements (symmetry pairs), as well as single elements that constitute a symmetrical configuration.

We consider *compositional structure* **(C1)** of images and *similarity* **(C2)** of constituent elements, in particular perceptual similarity in the low-level features and semantic similarity of objects and regions. The resulting model of symmetrical structure consists of a set of image elements and the pairwise similarity relations between the elements.

*(***C1***) Compositional Structure*

Symmetrical composition in the case of reflection consists of pairs of image elements in which one component is on the left and another is on the right of the symmetry axis, and in which single centered image elements are placed on the axis. To model this, we represent the extracted image elements as spatial entities, such as *points*, *axis-aligned rectangles*, and *line segments*, and we define constraints on the spatial configuration of those elements using five *spatial properties*:

- *position:* the center point of a rectangle or position of a point in x, y coordinates;
- *size:* the width and height of a rectangle w, h;
- *aspect ratio:* the ratio r between width and height of a rectangle;
- *distance:* euclidian distance d between two points p and q;
- *rotation:* the $yaw, pitch$, and $roll$ angles between two line segments in 3D space.

Symmetrical Spatial Configuration. We use a set of spatial relations that hold between the image elements to express their spatial configuration. Spatial relations (e.g., $left$, $right$, and on)[1] that hold between points and lines describe the relative orientation of image elements with respect to the symmetry axis. For this purpose, we use the relative position ($\mathrm{rel\text{-}pos}$) of an image element with respect to the symmetry axis, which is the distance to the symmetry axis and the element's y coordinate.

Image Patches and Objects. Symmetrical configuration of image elements is defined in terms of spatial properties using two rules. For a single element e, the center of the rectangle must be on the symmetry axis:

$$\mathsf{symmetrical}(e) \supset \mathsf{orientation}(on, \mathsf{position}(e), \mathrm{symmetry\text{-}axis}). \tag{1}$$

Pairs of elements e_i and e_j must be on opposite sites of the symmetry axis, must have the same size and aspect ratio, and their positions must be reflected:

$$\begin{aligned} &\mathsf{symmetrical}(p_i, p_j) \supset \\ &\quad \mathsf{orientation}(left, \mathsf{position}(p_i), \mathrm{symmetry\text{-}axis}) \wedge \\ &\quad \mathsf{orientation}(right, \mathsf{position}(p_j), \mathrm{symmetry\text{-}axis}) \wedge \\ &\quad \mathsf{equal}(\mathsf{aspect\text{-}ratio}(p_i), \mathsf{aspect\text{-}ratio}(p_j)) \wedge \\ &\quad \mathsf{equal}(\mathsf{size}(p_i), \mathsf{size}(p_j)) \wedge \mathsf{equal}(\mathsf{rel\text{-}pos}(p_i), \mathsf{rel\text{-}pos}(p_j)). \end{aligned} \tag{2}$$

This definition of symmetry serves as a basis for analyzing structures and specifies the attributes that constitute a symmetrical configuration.

Human Body Pose. We also define rules for symmetry in the placement and layout of humans in images. Given a set of joints j represented as points, the pose $pose = \{j_0, ..., j_n\}$ can be either symmetrical by itself or two people can be arranged in a symmetrical way. Body pose is analyzed by defining joint pairs $JP = \{(j_k, j_l), ..., (j_m, j_n)\}$, such as $(left\ shoulder, right\ shoulder)$ and $(left\ elbow, right\ elbow)$, then comparing the relative position of these pairs with respect to the center of the person c_p:

$$\mathsf{symmetrical}(\mathsf{pose}(p)) \supset \forall (j_k, j_l)\ \mathsf{equal}(\mathsf{rel\text{-}pos}(j_k, c_p), \mathsf{rel\text{-}pos}(j_l, c_p)). \tag{3}$$

1. The semantics of spatial relations is based on specialized polynomial encoding, as suggested in Bhatt et al. (2011), within constraint logic programming (Jaffar & Maher, 1994). We also use this framework to demonstrate question answering later in the section.

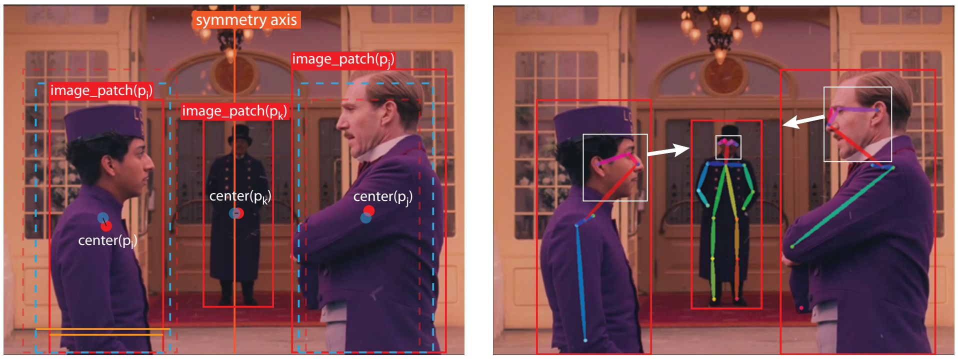


Figure 4. Symmetric composition for pairs of image patches, and centering of single image patches.

Accordingly, the pose of two persons is analyzed by defining joint pairs that associate each joint of one person to the corresponding joint of the other person, such as associating the left hand of person 1 with the right hand of person 2.

Further, we define symmetrical facing directions based on the rotation of their heads. Here we use the $yaw, pitch$, and $roll$ angles of a person's head h_p, relative to a front-facing head, and say the facing direction is symmetrical if the $pitch$ rotation is the same and if the yaw and $roll$ rotations are opposite:

$$\begin{aligned}&\mathsf{symmetrical}(\mathsf{facing\_dir}(p_1), \mathsf{facing\_dir}(p_2)) \supset \\ &\quad \mathsf{equal}(\mathsf{pitch}(h_{p_1}), \mathsf{pitch}(h_{p_2})) \wedge \mathsf{equal}(\mathsf{yaw}(h_{p_1}), -\mathsf{yaw}(h_{p_2}) \wedge \mathsf{equal}(\mathsf{roll}(h_{p_1}), -\mathsf{roll}(h_{p_2})).\end{aligned} \tag{4}$$

Divergence from Symmetrical Configuration. To account for configurations that only approximate symmetry, as typically occurs in naturalistic scenes, we calculate the divergences of the configuration from the symmetry model. For each element of the structure, we calculate the divergence from ideal symmetry in terms of position, size, aspect ration, and pose of body parts and joints. We use thresholds on the average of these values to identify hypotheses on (a)symmetrical structures.

*(**C2**) Similarity Measures*

Visual Symmetry is also based on similarity of image features. We assess similarity of image patches using features in a convolutional neural network, such as that obtained from AlexNets (Krizhevsky et al., 2012) or ResNets (He et al., 2016) when pre-trained on the ImageNet data set (Deng et al., 2009). We use the extracted features to evaluate perceptual similarity and use ImageNet classifications to evaluate semantic similarity of image patches.

Perceptual Similarity. Visual Symmetry is based in perceptual similarity of features extracted from the appearance of image patches. To analyze this perceptual similarity, we use cosine similarity over the feature vectors obtained from the network for two image patches. For reflectional symmetry, we examine the image patches for all potential symmetry pairs by comparing the features of one patch to the features of the second patch.

Table 1. Sample predicates for querying interpretation model.

Predicate	**Description**
symmetrical_element(E)	Symmetrical elements E.
non_symmmetrical_element(E)	Non-symmetrical elements E.
symmetrical_objects(SO)	Symmetrical objects SO.
non_symmetrical_objects(NSO)	Non-symmetrical objects NSO.
symmetrical_body_pose(SP,SBP)	Symmetrical person SP (pair or single object), and symmetrical parts of body-pose SBP.
non_symmetrical_body_pose(SE,NSP)	Symmetrical person SP (pair or single object), and non-symmetrical parts of body-pose SBP.
symmetry_stats(NP, NSP, MD, MS)	Basic stats on symmetrical structure: number of patches NP, number of symmetrical patches NSP, mean divergence MD, and mean similarity MS.
symmetrical_objects_stats(NO, NSO, MD, MS)	Stats on symmetrical structure of objects: number of objects NO, number of symmetrical objects NSO, mean divergence MD, and mean similarity MS.

Semantic Similarity. On the semantical level, we classify the image patches and compare their categories for conceptual similarity. For this, we use the weighted ImageNet classifications for each image patch with WordNet (Miller, 1995), which underlies the structure of ImageNet, to estimate conceptual similarity of object classes predicted for image patches in each symmetry pair. In particular, we use the top five predictions from AlexNet classifiers and estimate similarity of each pair as the weighted sum of similarity values for each pair of predicted categories.

2.3 Semantics of Declarative Symmetry

The semantic structure of symmetry is described by the model in terms of a set of symmetry pairs and their respective similarity values with respect to the three layers of our model. For each symmetry pair, it provides measures based on semantic similarity, spatial arrangement, and low-level perceptual similarity, as shown in Figure 5. This provides a declarative model of symmetrical structure that is used for fine-grained analysis of features and question-answering about configuration in images. We can use this framework to define high-level rules and execute queries in constraint logic programming (Jaffar & Maher, 1994) using SWI-Prolog (Wielemaker et al., 2012) to reason about symmetry and to directly *query* symmetrical features of the image.[2]

Symmetrical Structure of Images. As an example, consider the image in Figure 5. Based on the extracted symmetrical structure, the underlying interpretation model can be queried using utility predicates such as those in Table 1. The symmetry model as defined in Section 2.2 can be used to query symmetrical and non-symmetrical elements of the image using two rules:

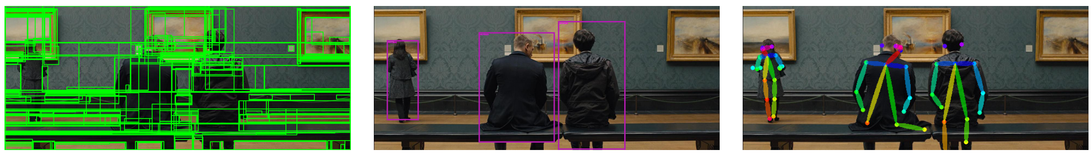
2. Within the constraint logic programming language PROLOG, ‘ **,** ’ corresponds to conjunction, ‘ **;** ’ to a disjunction, and ‘a **:-** b, c.’ denotes a *rule* that states ‘a’ is true if both ‘b’ and ‘c’ are true. Capital letters denote variables, whereas lower-case letters refer to constants, and ‘_’ (i.e., the underscore) is a “wild card” variable for cases in which one does not require a resulting value.

Step 1) *Extracting Image Elements.* Extract image elements $\mathcal{E}$ consisting of $\mathcal{E}_1, \mathcal{E}_2, and\ \mathcal{E}_3$:

($\mathcal{E}_1$) *Image Patches* are extracted using selective search as described in Uijlings et al. (2013);

($\mathcal{E}_2$) *People and Objects* are detected in the image using YOLO object detection (Redmon et al., 2016);

($\mathcal{E}_3$) *Human Body Pose* consisting of body joints and facing direction is extracted using human pose estimation (Cao et al., 2017).

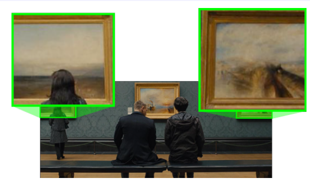


```
patch(id(0), rectangle(point(233, 53), 107, 466) ).
object(id(0), type('person'),rectangle(point(392,106),261,381)).
person(id(0), joint(id(0)), point((582, 159))).
...
```

Step 2) *Semantic and Perceptual Similarity.* Compute semantic and perceptual similarity for each pair of image elements e_i and $e_j \in \mathcal{E}$ based on features from CNN layers.

– Compute *semantic similarity* based on ImageNet classification of image patches

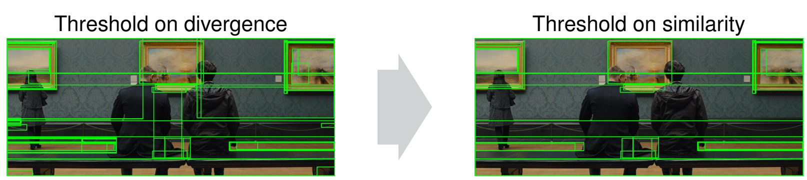
– Compute *perceptual similarity* based on cosign similarity between CNN features of image elements



```
category(img_patch(170),
        [category("window shade", 0.456657), ..., category("television", 0.0901952)]).
category(img_patch(200),
        [category("shoji", 0.455987), ..., category("window shade", 0.0961778)]).
...
similarity(perceptual, pair(170, 200), 0.70728711298).
similarity(semantic, pair(170, 200), 0.6666666666666666).
...
```

Step 3) *Symmetry Configuration and Divergence.* Identify symmetrical structures in the image elements $\mathcal{E}$ based on the formal definition of symmetry in Section 2.2 and calculate the divergence of elements from this model.

```
?- divergence(symmetrical(id(170), id(200)), DivSize, DivAR, DivPos).
DivSize = div_size(9.0, 18.0),  DivAR = div_ar(0.0595),  DivPos = div_pos(3.9051)
```



Result *A Declarative Model for Semantic Analysis of Symmetry.* The process results in the declarative description of the symmetric properties in the image given by the image elements $\mathcal{E}$ and the divergence from the formal definition of symmetry. This model serves as a basis for declarative query answering about symmetrical characteristics of the image. For example, the images showcase results from queries in Section 2.3, analyzing symmetrical and non-symmetrical elements.

Figure 5. Computational steps to generate the semantic symmetry model.

```
symmetrical_element(E) :- symmetrical(E).
symmetrical_element(E) :- symmetrical(E, _); symmetrical(_, E).
```

Aggregating results for the symmetrical_element(E) predicate for the example image results in a list of all symmetrical image elements, as Figure 5 depicts:

```
SYMETRICAL = [0, 2, 8, 10, 11, 12, 14, 15, 17|...]
```

Similarly, we can query the non-symmetrical elements of the image using the rule:

```
non_symmetrical_element(P) :- image_element(P), not(symetrical_element(P)).
```

```
NON_SYMETRICAL = [1, 3, 4, 5, 6, 7, 9, 13, 16|...].
```

Divergence. The divergence of image elements from the ideal symmetrical configuration can be directly queried using the divergence predicate:

```
?- divergence(symmetrical(id(P1), id(P2)), Div_Size, Div_AR, Div_Pos).
...
P1 = 170, P2 = 200,
DivSize = div_size(9.0, 18.0),
DivAR = div_ar(0.0595206914614983),
DivPos = div_pos(3.905124837953327);
...
```

Similarity. Perceptual and semantic similarity of image elements are queried as:

```
?- similarity(pair(id(P1), id(P2)), Percept_Sim, Semantic_Sim).
...
P1 = 170, P2 = 200,
Percept_Sim = 0.70728711298,
Semantic_Sim = 0.6666666666666666;
...
```

These predicates provide the basis for the semantic analysis of symmetry structures in the image that we describe later.

Symmetrical Structure of Objects and People. Symmetrical structures in configurations of *objects* and *people* in an image can be queried using the predicate symmetrical_objects to get pairs of symmetrically positioned objects and single objects that appear in the center of the image.

```
?- symmetrical_objects(SymObj).
```

For the example image this results in the two people sitting on the bench in the center of the image.

```
SymObj = pair(id(1), id(2)).
```

Similarly to symmetrical object configurations, objects placed in a *non-symmetrical* way can be queried as:

```
?- non_symmetrical_objects(NonSymObj).
```

```
NonSymObj = id(0).
```

This returns objects that are not part of a symmetrical structure, in that the person on the left of the image has no symmetrical correspondent on the right.

Body Pose. Based on this, the extracted symmetrical objects can be analyzed further, to query the symmetrical configuration of *people* and their *body poses*:

```
symmetrical_body_pose(pair(P,Q), SymPose) :-
        symmetrical_objects(pair(P, Q)),
        type(P, 'person'), type(Q, 'person'),
        symmetrical(pose(pair(P, Q)), SymPose).
```

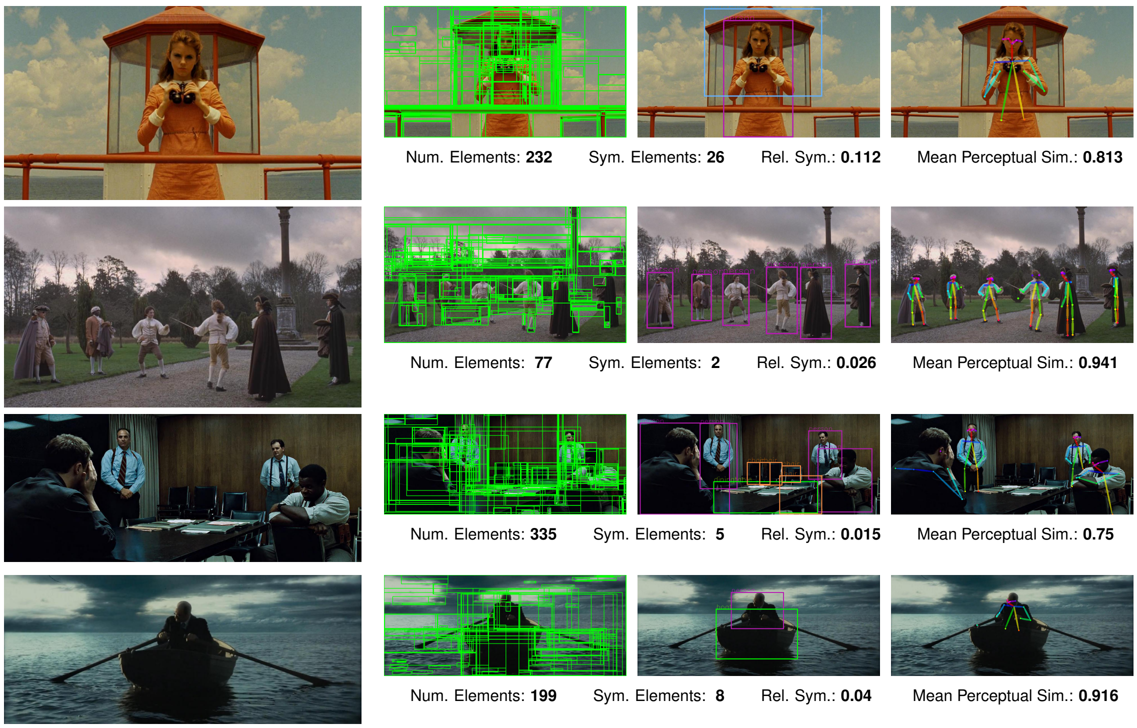


Figure 6. Extracted elements and statistics on symmetry structures for exemplary images.

This produces the symmetrically placed people and the elements of their poses that are symmetrical. In this case, the upper-body of person 1 and person 2 satisfy the definition:

```
P = id(1), Q = id(2),
SymPose =   ['upperbody'].
```

Respectively, we can query *non-symmetrical parts* of the *body pose*:

```
?- non_symmetrical_body_pose(pair(P, Q), NonSymPose).
```

This gives those parts of the body poses that diverge from symmetry:

```
P = id(1), Q = id(2),
NonSymPose =  ['facing direction', 'legs'].
```

The above analysis reveals that the two people are sitting on the bench in a symmetrical way. Their upper-body pose is symmetrical, while the facing direction and the legs are not symmetrical.

Statistics on Image Symmetry. Additionally, we can use the model to query statistics on the symmetrical features of an image, which in turn we can use to train a classifier based on the semantic characterizations of symmetry from Section 3. Figure 6 presents additional examples of such statistics.

```
P = id(1), Q = id(2),
NonSymPose =  ['facing direction', 'legs'].
```

Once they have been computed, we can query statistics on the symmetry of objects and people:

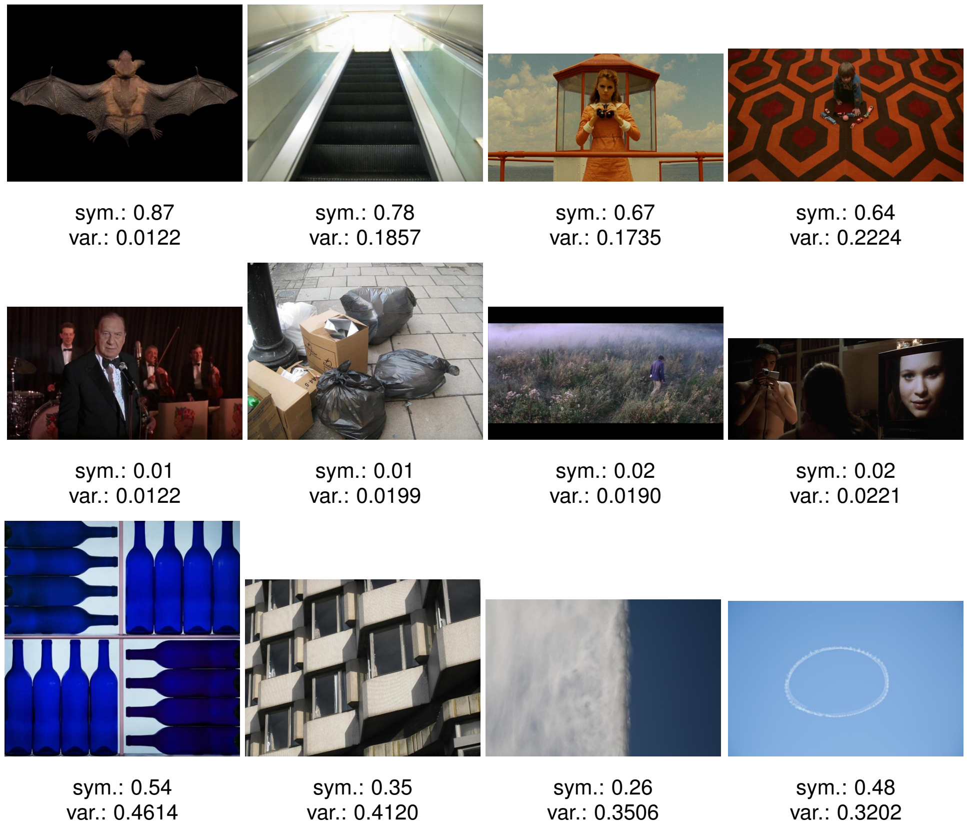


Figure 7. Sample results from the human experiment. (**row 1**) most symmetric; (**row 2**) most non-symmetric (these correspond directly to the images with the lowest variance in the answers); (**row 3**) images with the biggest variance in the answers.

```
?- symmetry_stats(NumImgPatches, NumSymPatches, MeanDiv, MeanSim).
NumImgPatches = 359,
NumSymPatches =  40,
MeanDiv = [div_w(12.394), div_h(7.394), div_ar(0.944), div_pos(8.32)],
MeanSim = 0.8162167312386968.
```

In summary, these rules provide a declarative, interpretable characterization of reflectional symmetry as it arises in visual stimuli.

3. Human Evaluation: A Qualitative Study

Human-generated data from subjective, qualitative assessments of symmetry serves many useful purposes, so we built a data set of 150 images that consisted of landscape, architectural, and movie scenes. The images ranged from highly symmetric ones with symmetric patterns to completely non-symmetric images. We showed each participant 50 images selected randomly from the data set; subjects were to rank the pictures by selecting one of four categories: not_symmetric, some-

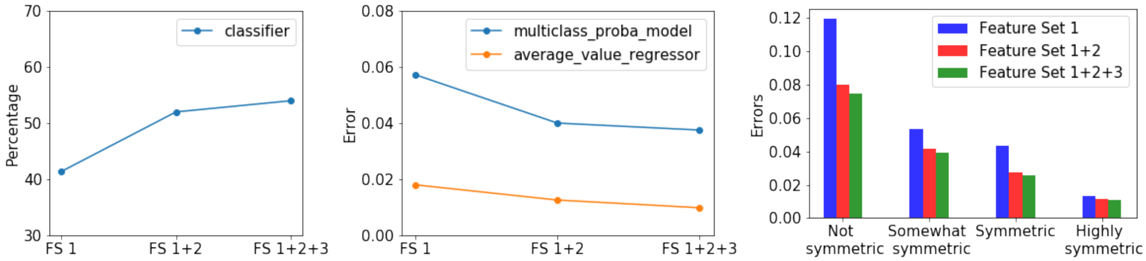
(a) (b) (c)

Figure 8. Samples from the experimental data.

what_symmetric, symmetric, and highly_symmetric. Each image was presented to approximately 100 participants. We calculated the symmetry value as the average of all responses.

The results from this experiment suggest that perception of symmetry varies substantially across subjects. People tend to agree about when there is no symmetry, with variance in answers being very low, as Figure 7 shows. For high-symmetry cases, there is a greater variation in the human perception. For images with an average level of symmetry, the variance is especially high. Qualitatively, we found three results. (1) *Absence of features* decreases the subjective rating of symmetry: the image in Figure 8a has nearly perfect symmetry but, as it has few features that can be symmetrical, people only perceived it as having medium symmetricality, with a high variance in the answers. (2) *Symmetrical placement of people* in the image has a higher impact on the subjective judgement of symmetry than other objects: the image in Figure 8b is judged as symmetrical based on the placement of the characters and the door in the middle, but the objects on the left and right side are not very symmetrical. (3) Images that are *naturally structured* in a symmetrical way are judged less symmetrical than those with objects arranged in a symmetrical way: images of centered faces depicted in Figure 8c are rated less symmetrical than other images with similar symmetry on the feature level.

To evaluate how well our symmetry model reflects these human criteria for judging symmetry in naturalistic images, we use the results from the human study to train a classifier and a regressor to predict the symmetry class of an image and predict the average level of perceived symmetry of the images. We extracted three sets of features ($fs1$–$fs3$) from the symmetry model: $fs1$ consists of the cosine similarity between the two halves of each image on the five convolution layers in an AlexNet; $fs2$ consists of the symmetric properties between image patches, measured by their divergence from symmetrical spatial configuration, and perceptual similarity; and $fs3$ consists of the symmetric properties of subjects and their configurations in the images. We have two models, a classifier and a regressor. A given image is classified into one of the four symmetry classes using the classifier. This model is evaluated using the mean accuracy, as shown in Figure 9(a). The classifier model also predicts the per class probabilities, which is denoted by $multiclass\_proba\_model$. This model is evaluated by calculating the mean squared error (MSE) between the predicted probabilities and the percentages from the human data for each class. The per class errors are shown in Figure 9(c), while the mean error is shown in Figure 9(b). The regressor model predicts the average symmetry value of a given image. The model is evaluated by calculating the MSE between the predicted average symmetry value and average symmetry value from the human data. We use the pipeline optimization method of TPOT (Olson et al., 2016) to automatically build the classification



(a) (b) (c)

Figure 9. Results of empirical evaluation with three different feature set combinations, showing (a) mean accuracy, (b) mean error, and (c) class probability error.

and regression pipelines for the feature sets. This results in a classification pipeline consisting of an ensemble of decision trees, support vector machines, and random forest classifiers while the regression pipeline consists of an ensemble of ExtraTrees and XGBoost regressors. The models are trained and tested on the three-feature set using cross validation, splitting the 150 images into five folds. Reported are *mean error* and *classification accuracy* (CA).

The results in Figure 9 and Table 2 show that using the features from our symmetry model improves performance in both tasks, i.e., the accuracy for the classification task improves by over 10% (Table 2) from 41.33 % to 54%, and the error for per class probabilities decreases from 0.057 to 0.038. The biggest improvement in the classification and in the prediction of the average symmetry value happens when adding the image patch features $fs2$, Figure 9(a) and (b). Adding people-centered features only results in a small improvement, which may be because only a subset of the images in the data set involves people. The results on the predicted per class probabilities, seen in Figure 9(c), show that adding features from our symmetry model leads to better predictions of the variances in the human answers.

4. Related Research

Symmetry in images has been studied from many different perspectives, including visual perception research, neuroscience, cognitive science, arts, and aesthetics (Treder, 2010). The semantic interpretation of symmetry in perception and aesthetics requires a mixed methodology consisting of both empirical and analytical methods:

- **Empirical** / *Human behavior Studies*. This involves qualitative studies involving subjective assessments, as well as an evidence-based approach measuring human performance from the viewpoint of visual perception using eye-tracking, qualitative evaluations, and think-aloud analysis with human subjects.
- **Analytical** / *Interpretation and Saliency*. This involves the development of computational models that serve an interpretation and a predictive function involving, for instance: (i) multi-level computational modelling of interpreting visuo-spatial symmetry; (ii) a saliency model of visual attention serving a predictive purpose vis-a-vis the visuo-spatial structure of visual media.

Table 2. Results from classification and prediction pipeline.

Feature Sets	CA (%)	Avg. Sym. Err.	Class Prob. Err.
$fs1$	41.33	0.01806876383	0.0572886659
$fs1$+2	52.00	0.0126452444	0.0400713172
$fs1$+2+3	54.00	0.009900461023	0.0375853705

There are numerous studies investigating how symmetry affects visual perception (Cohen & Zaidi, 2013; Norcia et al., 2002; Machilsen et al., 2009; Bertamini & Makin, 2014), and how it is detected by humans (Wagemans, 1997; Freyd & Tversky, 1984; Árpád Csathó et al., 2004). Most relevant to our work is the research on computational symmetry in the area of computer vision (Liu et al., 2013, 2010). Typically, computational studies on symmetry characterize symmetry in reflection, translation, or rotation, with the first (also referred to as *bilateral* or *mirror symmetry*) most extensively investigated. Another direction of research focuses on detecting symmetric structures in objects. Teo et al. (2015) described a classifier that detects curved symmetries in 2D images, and Lee and Liu (2012) presented an approach to detect curved glide-reflection symmetry in 2D and 3D images. Atadjanov and Lee (2016) used the appearance of structural features to detect symmetry in objects.

More generally, analyzing image structure is a central topic in computer vision and there are a variety of methods proposed for this task. Convolutional neural networks provide a basis for analyzing images using learned features, with AlexNets (Krizhevsky et al., 2012) and ResNets (He et al., 2016), trained on the ImageNet data set (Deng et al., 2009), being good examples. Recent developments in object detection extend the concept to detection with 'region' convolutional neural networks (Girshick et al., 2016; Ren et al., 2017). In these methods, objects are detected based on region proposals extracted from the image, for example, using selective search (Uijlings et al., 2013) or region proposal networks for predicting object regions. Finally, for image comparison, Zagoruyko and Komodakis (2015) and Dosovitskiy and Brox (2016) measure perceptual similarity based on features learned by a neural network.

5. Conclusions

Our research addresses visuo-spatial symmetry in the context of naturalistic stimuli from visual arts, including film, paintings, and photography. With a principal focus on developing a computational model of interpreting visuo-spatial symmetry in a human-like manner, our approach is driven by three mutually synergistic aspects, namely, reception, interpretation, and synthesis:

- **Reception**: We performed a behavioral study on human perception and explanation of symmetry, with emphases on visual attention and spatio-linguistic and qualitative characterizations;
- **Interpretation**: We developed a computational model that can interpret deep semantic meanings of visual symmetry with human-like explainability and visual sense-making;
- **Synthesis**: The system is capable of applying human-like explanation models as a basis of either directly or indirectly engineering visual media vis-a-vis their predictive receptive effects, namely, guiding attention by influencing visual fixation patterns, or manipulating saccadic movements in a variety of situations.

Figure 10. Manual manipulation of symmetry. Symmetry decreasing from highly symmetric to not symmetric.

In this paper, we focused on the reception and interpretation aspects. We presented a computational model of reflectional symmetry that integrates visuospatial composition, feature-level similarity, and semantic similarity in visual stimuli. Based on this result, some possible next steps include:

- **Spatio-temporal symmetry**. On the computational front, we plan to extend the symmetry model beyond static images to analyze symmetry in space-time image segments (e.g., in Wes Anderson films – Bhatt & Suchan, 2015), animation, and other kinds of narrative media. This will involve incorporating a richer spatio-temporal ontology (Suchan & Bhatt, 2016b; Suchan et al., 2018; Schultz et al., 2018) and a particular focus will be the influence of space-time symmetry on visual fixations and saccadic eye-movements (Suchan et al., 2016b).
- **Visual processing aspect**. We will try more advanced region proposals that use different visual computing primitives and similarity measures.
- **Resynthesizing images**. Using our framework, we plan to produce qualitatively distinct classes of (a)symmetry like those shown in Figure 10, and conduct further empirical studies.

Additionally, we would welcome work in complementary areas. We envisage these including methods to provide a holistic view of the cinematographic "**geometry of a scene**" (Suchan & Bhatt, 2016a,b), relational learning of visuo-spatial symmetry patterns, such as those based on inductive generalization (Suchan et al., 2016a)), and explainable learning from visual data sets for a new approach to the study of media and art history, cultural studies, and aesthetics.

Acknowledgements

We thank the anonymous reviewers and the volume editors for constructive comments that helped improve the paper. The empirical aspect of this research was made possible by the generous participation of volunteers in the online survey, and we are grateful to all the participants. We acknowledge the support of Steven Kowalzik for manual resynthesis of stimuli categories.

References

Atadjanov, I. R., & Lee, S. (2016). Reflection symmetry detection via appearance of structure descriptor. *Proceedings of the Fourteenth European Conference on Computer Vision* (pp. 3–18). Amsterdam, The Netherlands: Springer.

Bertamini, M., & Makin, A. D. (2014). Brain activity in response to visual symmetry. *Symmetry*, *6*, 975–996.

Bhatt, M., Lee, J. H., & Schultz, C. P. L. (2011). CLP(QS): A declarative spatial reasoning framework. *Proceedings of the Tenth International Conference on Spatial Information Theory* (pp. 210–230). Belfast, ME: Springer.

Bhatt, M., Schultz, C., & Freksa, C. (2013). The 'space' in spatial assistance systems: Conception, formalization and computation. In T. Tenbrink, J. Wiener, & C. Claramunt, (Eds.), *Representing space in cognition: Interrelations of behavior, language, and formal models.* Oxford, UK: Oxford University Press.

Bhatt, M., & Suchan, J. (2015). Obsessed by symmetry: Wes Anderson and the visual and cinematographic shape of his films. *SHAPES 3.0: Proceedings of the Third Interdisciplinary Workshop.* Larna, Cyprus.

Cao, Z., Simon, T., Wei, S., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1302–1310). Honolulu, HI: IEEE Press.

Cohen, E. H., & Zaidi, Q. (2013). Symmetry in context: Salience of mirror symmetry in natural patterns. *Journal of Vision*, *13*, 22.

Árpád Csathó, van der Vloed, G., & van der Helm, P. A. (2004). The force of symmetry revisited: Symmetry-to-noise ratios regulate (a)symmetry effects. *Acta Psychologica*, *117*, 233–250.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition.* (pp. 248–255). Miami, FL: IEEE Press.

Dosovitskiy, A., & Brox, T. (2016). Generating images with perceptual similarity metrics based on deep networks. *Proceedings of the Thirtieth Advances in Neural Information Processing Systems*. Barcelona, Spain: NIPS.

Freyd, J., & Tversky, B. (1984). Force of symmetry in form perception. *The American Journal of Psychology*, *97*, 109–126.

Girshick, R. B., Donahue, J., Darrell, T., & Malik, J. (2016). Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*, 142–158.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). Las Vegas, NV: IEEE Press.

Jaffar, J., & Maher, M. J. (1994). Constraint logic programming: A survey. *The Journal of Logic Programming*, *19*, 503–581.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Proceedings of the Twenty-Sixth Advances in Neural Information Processing Systems* (pp. 1097–1105). Lake Tahoe, NV: NIPS.

Lee, S., & Liu, Y. (2012). Curved glide-reflection symmetry detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*, 266–278.

Liu, J., Slota, G., Zheng, G., Wu, Z., Park, M., Lee, S., Rauschert, I., & Liu, Y. (2013). Symmetry detection from realworld images competition 2013: Summary and results. *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Portland, OR: IEEE Press.

Liu, Y., Hel-Or, H., Kaplan, C. S., & Gool, L. V. (2010). Computational symmetry in computer vision and computer graphics. *Foundations and Trends in Computer Graphics and Vision*, *5*, 1–195.

Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, *9*, 11.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*, 39–41.

Molnar, V., & Molnar, F. (1986). Symmetry-making and -breaking in visual art. *Computers & Mathematics with Applications*, *12*, 291–301.

Norcia, A. M., Candy, T. R., Pettet, M. W., Vildavski, V. Y., & Tyler, C. W. (2002). Temporal dynamics of the human response to symmetry. *Journal of Vision*, *2*, 132–139.

Olson, R. S., Urbanowicz, R. J., Andrews, P. C., Lavender, N. A., Kidd, L. C., & Moore, J. H. (2016). Automating biomedical data science through tree-based pipeline optimization. *Proceedings of the Nineteenth European Conference on Applications of Evolutionary Computation* (pp. 123–137). Porto, Portugal: Springer.

Redmon, J., Divvala, S. K., Girshick, R. B., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788). Las Vegas, NV: IEEE Press.

Ren, S., He, K., Girshick, R. B., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*, 1137–1149.

Schultz, C., Bhatt, M., Suchan, J., & Wałęga, P. (2018). Answer set programming modulo 'space-time'. *Proceedings of the Second International Joint Conference on Rules and Reasoning* (pp. 1–16). Luxembourg: Springer.

Suchan, J., & Bhatt, M. (2016a). The geometry of a scene: On deep semantics for visual perception driven cognitive film, studies. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision* (pp. 1–9). Lake Placid, NY: IEEE Press.

Suchan, J., & Bhatt, M. (2016b). Semantic question-answering with video and eye-tracking data: AI foundations for human visual perception driven cognitive film studies. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (pp. 2633–2639). New York: AAAI Press.

Suchan, J., Bhatt, M., & Schultz, C. P. L. (2016a). Deeply semantic inductive spatio-temporal learning. *Proceedings of the Twenty-Sixth International Conference on Inductive Logic Programming* (pp. 73–80). London, UK.

Suchan, J., Bhatt, M., Wałęga, P. A., & Schultz, C. P. L. (2018). Visual explanation by high-level abduction: On answer-set programming driven reasoning about moving objects. *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence* (pp. 1–12). New Orleans, LA: AAAI Press.

Suchan, J., Bhatt, M., & Yu, S. (2016b). The perception of symmetry in the moving image: Multi-level computational analysis of cinematographic scene structure and its visual reception. *Proceedings of the ACM Symposium on Applied Perception* (p. 142). Anaheim, CA: ACM.

Teo, C. L., Fermüller, C., & Aloimonos, Y. (2015). Detection and segmentation of 2D curved reflection symmetric structures. *Proceedings of the 2015 IEEE International Conference on Computer Vision* (pp. 1644–1652). Santiago, Chile: IEEE Press.

Treder, M. S. (2010). Behind the looking-glass: A review on human symmetry perception. *Symmetry*, *2*, 1510–1543.

Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T., & Smeulders, A. W. M. (2013). Selective search for object recognition. *International Journal of Computer Vision*, *104*, 154–171.

Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences*, *1*, 346–352.

Weyl, H. (1952). *Symmetry*. Princeton, NJ: Princeton University Press.

Wielemaker, J., Schrijvers, T., Triska, M., & Lager, T. (2012). SWI-Prolog. *Theory and Practice of Logic Programming*, *12*, 67–96.

Zagoruyko, S., & Komodakis, N. (2015). Learning to compare image patches via convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4353–4361). Boston, MA: IEEE Press.