

Global Semantic Description of Objects based on Prototype Theory

Omar Vidal Pino · Erickson R. Nascimento · Mario F. M. Campos

Received: date / Accepted: date

Abstract In this paper, we introduce a novel semantic description approach based on Prototype Theory foundations. Inspired by the human approach used for representing categories, we propose a Computational Prototype Model (CPM) that *encodes* and *stores* the central semantic meaning of the object's category: the semantic prototype. Also, we introduce a Prototype-based Description Model that encodes the semantic of an object while describing its features using our CPM model. Our description method uses semantic prototypes computed by convolutional neural network (CNN) classification models to create discriminative signatures that describe an object highlighting its most distinctive features within the category. Our experiments show that: *i*) the proposed CPM model (semantic prototype + distance metric) successfully describes the internal semantic structure of objects categories; *ii*) our semantic distance metric can be understood as object visual typicality score within a category; and *iii*) our descriptor encoding is semantically interpretable and significantly outperforms other image global encodings in clustering and classification tasks.

Keywords Semantic representation · Category representation · Object semantic features · Global features description · Prototype Theory

Omar Vidal Pino · Erickson R. Nascimento · Mario F. M. Campos
Universidade Federal de Minas Gerais, Computer Science Department, Computer Vision and Robotics Lab, Belo Horizonte, Minas Gerais, Brazil
Tel.: +55-31-3409 5856
Fax: +55-31-3409 5858
E-mail: {ovidalp, erickson, mario}@dcc.ufmg.br.

1 Introduction

Memory is one of the most amazing faculties of the human being. It is generally considered as the brain ability to *code*, *store*, and *retrieve* information (Atkinson and Shiffrin 1968). *Semantic memory* (Tulving 2007), for instance, refers to general world knowledge that we accumulate throughout our lives (McRae and Jones 2013). A relevant aspect of the functional neuroanatomy of the semantic memory resides in the *representation of the meaning* of objects and their properties (Martin 2007). Several assumptions indicate that human beings are capable of: *i*) learning the most distinctive features of a specific object category (Martin 2007; Thompson-Schill 2003); *ii*) form categories and *object semantic definitions* (abstractions) at a very early age (Martin 2007). Semantic memory involves the semantic definition of objects (Tulving 2007) and, consequently, the success of object recognition, classification, and description tasks are causally related to the success of effectively recovering the learned knowledge (Tulving 2007).

For several years, the fields of Computer Vision and Machine Learning have tried to build and learning pattern recognition methods with a similar performance of a human being for visual information processing. Although the state-of-the-art methods have achieved surprising results, there are still many challenges to achieve the discriminative power and abstraction of semantic memory to represent the semantic. How to describe and stand for objects, semantically? How to simulate the behavior of semantic memory in the representation of learned knowledge of objects' features? How to extract and encode the object features to encapsulate the representation of the meaning (or *semantic representation*) of a specific object? How to learn the semantic definition of categories objects and use this definition in

object recognition, classification, and description tasks? These are just some of the interesting questions that still occupy the investigation agenda of many research areas.

In this paper – motivated by the semantic memory behavior – we propose a mathematical model that attempts to represent the semantic definition of object categories. Also, we propose a procedure to introduce this semantic representation of object categories in the global description of objects features extracted from images.

The knowledge extraction models (high-level vision processes) from images are highly influenced by the methods used for detection, extraction, and representation of image relevant information. Consequently, the extraction of image relevant features has been the subject of Computer Vision research for decades. For several years, hand-crafted features (Bay et al. 2008; Lowe 2004; Tola et al. 2008) and machine learning methods (Simonyan et al. 2014; Strecha et al. 2012) were the choice for image feature description tasks.

The advent of Convolutional Neural Networks (CNN) outperformed these traditional methods and enabled them to achieve a visual recognition model with similar behavior of *semantic memory* for classification tasks (He et al. 2016; Simonyan and Zisserman 2014; Szegedy et al. 2017), sparking the tendency of images semantic processing with deep-learning techniques. The CNN-models success spawned numerous CNN-descriptors produced with different approaches that learn effective representations for describing image features (Han et al. 2015; Kim et al. 2018; Simo-Serra et al. 2015; Zagoruyko and Komodakis 2015). Consequently, representations of image features extracted using deep classification models (He et al. 2016; Simonyan and Zisserman 2014; Szegedy et al. 2017), or using CNN-descriptors are commonly referred as *semantic feature* or *semantic signature*.

Semantic feature term has been extensively studied in the field of linguistic semantic; it is defined as the representation of the basic conceptual components of the meaning of any lexical item (Fromkin et al. 2018). In the seminal work of Rosch (1975), the author analyzed the semantic structure of the meaning of words and introduced the concept of *semantic prototype*. According to Rosch (1975); Rosch and Mervis (1975), the representation of *category semantic meaning* is related to the *category prototype*, particularly to those categories naming natural objects.

Image semantic understanding is influenced by how are semantically represented the features of image basic components (*e.g.*, objects), and the semantic relations between these basic components (Guo et al. 2016).

CNN-description models (Han et al. 2015; Lin et al. 2016; Simo-Serra et al. 2015; Zagoruyko and Komodakis 2015) and semantic description models (Han et al. 2017; Kim et al. 2018; Rocco et al. 2018) stand for the semantic information of image features using a range of different approaches. Nevertheless, *none* of these models codify the representation of the visual information based on the theoretical foundation of Cognitive Science to represent the *semantic meaning*.

In this paper, we rely on cognitive semantic studies related to the Prototype Theory for modeling the *central semantic meaning* of objects categories: the prototype. We propose a novel approach to take on the semantic features descriptions of objects based on prototypes. Our *prototype-based description model* uses the category’s prototype to find a global semantic representation of the basics conceptual components (objects) of the image meaning.

To achieve this goal, we bring to light the Prototype Theory as a theoretical foundation to represent the semantic meaning of the visual information accurately to represent — semantically — the basics components of the image: objects. The Prototype Theory proposes that human beings think a category in terms of abstract prototypes, defined by typical members of the category (Geeraerts 2010; Rosch 1975, 1978). This theory also exposes that successful execution of object recognition and description tasks in the human brain is inherently related to the learned prototype of the object category (Minda and Smith 2002; Rosch 1975, 1978; Zaki et al. 2003). The observations on the Prototype Theory raise the following two questions: i) Can a model of the perception system be developed in which objects are described using the same semantic features that are learned to identify and classify them? ii) How can the category prototype be included in the object global semantic description?

We address these two questions motivated by the human’s approach to describing objects globally. Human being uses the generalization and discrimination processes to build object descriptions that highlighting their most distinctive features within the category. For example, a typical human description: a dalmatian is a dog (generalization ability to recognize the central semantic meaning of dog category) that is distinguished by its unique black, or liver-colored spotted coat (discrimination ability to detect the semantic distinctiveness of object within the dog category). Figure 1 illustrates the intuition and principal concepts of our prototype-based description model. The main idea of our approach is to use the quality of features extracted with CNN-classification models both to represent the

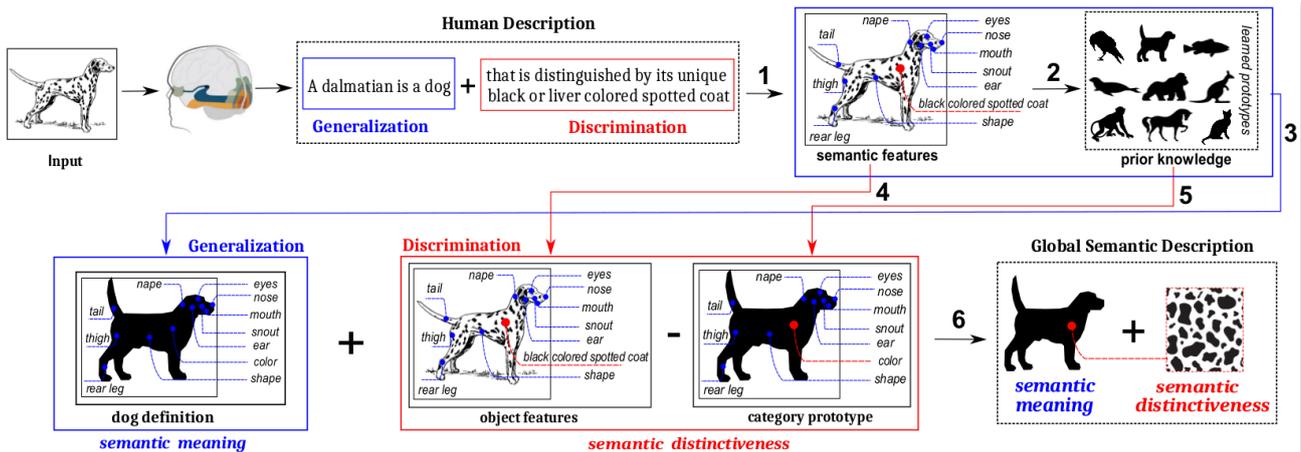


Fig. 1 *Motivation and Concepts.* Schematic of our prototype-based description model. The human visual system can observe an object and to build an object semantic description that highlighting their most distinctive features within the object category. We propose a prototype-based model to simulate this behavior through the processing flow from 1) to 6). 1) features extraction; 2) object features recognition; 3) categorization; 4) object features; 5) central semantic meaning of a category (the category prototype); 6) our Global Semantic Description based on Prototypes.

central semantic meaning of a specific category and learn the object distinctiveness within the category.

More specifically, our main contributions in this paper are as follows:

1. a *Computational Prototype Model* (CPM) based on Prototype Theory foundations, to stand for the central semantic meaning of object images categories. Our CPM model allows to interpret possible semantic associations between members within the category internal structure.
2. a *semantic distance metric* in object image CNN features domain, which can be understood as a measure of object typicality within the object category.
3. a *prototype-based description model* for global semantic description of objects images. Our semantic description model introduces, for the first time, the use of category prototypes in image global description tasks.

2 Related works

CNN descriptors

Descriptors extracted using CNN techniques have shown that it is possible, for a learning approach, to outperform the best techniques based on carefully hand-crafted features (Bay et al. 2008; Lowe 2004; Tola et al. 2008). CNN descriptor models differ among themselves on how to compute the descriptors in their deep architectures, similarity functions learning, and its features extraction methods. Some approaches extract immediate activations of the model as a descriptor signature (Si-

mony and Zisserman 2014; Szegedy et al. 2017; Donahue et al. 2014; Long et al. 2014). Others methods use similarity convolutional networks (Han et al. 2015; Simo-Serra et al. 2015; Yi et al. 2016; Zagoruyko and Komodakis 2015) and Siamese networks (Han et al. 2015; Zagoruyko and Komodakis 2015; Yi et al. 2016) to learn discriminative representations. LIFT (Yi et al. 2016) learns each task involved in features management: detection, orientation estimation, and description. Lin et al. (2016) constructed a compact binary descriptor for efficient object matching based on features extracted with VGG16 model (Simonyan and Zisserman 2014). Those CNN-descriptor models were more oriented to achieve discriminative features than representing the image semantic.

Semantic descriptors and semantic correspondence

Liu et al. (2011) proposed SIFT Flow method. SIFT Flow method generated the start of semantic flow family methods as a solution to the challenge of semantic correspondence (Bristow et al. 2015; Liu et al. 2011; Yang et al. 2014). Several of these methods combine their approaches with the extraction of hand-crafted features (Lowe 2004; Tola et al. 2008). Some works (Han et al. 2017; Kim et al. 2018; Rocco et al. 2018) use the robustness of CNN-models for training deep learning architectures and address the problem of semantic correspondence. Kim et al. (2018) tackled the problem of semantic correspondence by constructing FCSS semantic descriptor. In general, CNN descriptors and semantic descriptors are trained to learn their semantic repre-

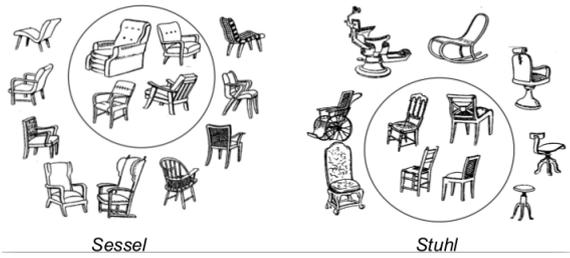


Fig. 2 *Category prototypicality organization.* Figure shows the *Sessel* and *Stuhl* experiment conducted by Gipper (Figure adapted from Geeraerts (2010)). That experiment studies the meaning of German words *Stuhl* (chair) and *Sessel* (comfortable chair) and shows that within the *chair* category, category *central semantic meaning* can change depending of observed feature relevance (weights) and object typicality. This phenomenon is described in contemporary semantics as a *prototypicality organization* (Rosch 1978; Geeraerts 2010) and constitutes one of the motivations of our proposal.

sentations and use different deep learning architectures. Most of these features description models do not use the discriminative power of the features extracted using the well-known CNN-classification models (He et al. 2016; Simonyan and Zisserman 2014; Szegedy et al. 2017). Moreover, none of these CNN-feature description approaches incorporates the foundation of the Cognitive Sciences to introduce *meaning* in the representations of image features.

Prototype Theory

The Prototype Theory (Rosch and Lloyd 1978; Rosch and Mervis 1975; Rosch 1988; Geeraerts 2010; Minda and Smith 2002; Rosch 1975, 1978; Zaki et al. 2003) analyzes the internal structure of categories and introduces the prototype-based concept of categorization. It proposes categories representation as heterogeneous and not discrete, where the features and category members do not have the same relevance within the category. Rosch (Rosch 1975, 1978) obtained evidence that human beings store first the *semantic meaning of category* based on the degrees of representativeness (*typicity*) of category members, and then its specificities.

The *category prototype* was formally defined as the clear central members of a category (Geeraerts 2010; Rosch 1975; Rosch and Mervis 1975). The attributes of these focal members are those that are structurally the most salient category properties, and conversely, a member occupies the focal position because it shows the most salient features of the category (Rosch and Mervis 1975; Geeraerts 2010). Rosch (Rosch 1975, 1978, 1988) showed that human beings store the category knowledge as a semantic organization around the cate-

Table 1 Two-dimensional conceptual map of prototypicality effects (Geeraerts 2010).

	<i>extensional</i>	<i>intensional</i>
<i>non-equality</i> (salience effect, core/periphery)	Difference of typicality and membership salience	Clustering into family resemblances
<i>non-discreteness</i> (demarcation problems, flexibly)	Fuzziness at the edges, membership uncertainty	Absence of necessary and sufficient definitions

gory prototype (*prototypicality organization*). Figure 2 shows an example of the *prototypicality organization* phenomenon (Rosch 1978; Rosch and Lloyd 1978; Geeraerts 2010). Finally, object categorization is obtained based on the similarity of a new exemplar with the learned categories prototypes (Rosch 1978, 1988).

Rosch (Rosch 1975, 1978; Rosch and Lloyd 1978; Rosch 1988) showed the important of making distinctions between various phenomena that may be associated with *prototypicality*. For Geeraerts (2010) the concept of prototypicality is in itself a prototypically clustered one for four characteristics in which the concepts of *non-discreteness* and *non-equality* (either on the *intensional* or on the *extensional* level) play a major distinctive role. Four characteristics are frequently mentioned as typical of prototypicality in prototypical categories (Rosch 1975, 1978; Geeraerts 2010): *i*) categories exhibit degrees of typicality; not every member is equally representative in the category (*extensional non-equality*); *ii*) categories are blurred at the edges (*extensional non-discreteness*); *iii*) categories are clustering into *family resemblance structure*; that is, the category semantic structure takes the form of a radial set of clustered and overlapping members (*intensional non-equality*); and *iv*) categories cannot be defined by means of a single set of criteria (necessary and sufficient) attributes (*intensional non-discreteness*). The *prototypicality effects* (see Table 1) surmise the importance of the distinction between central and peripheral meaning of the object categories (Geeraerts 2010).

3 Computational Prototype Model

Rosch (1975, 1978) showed that human beings learn the central semantic meaning of categories (the prototype) and include it in their cognitive processes. Based on these assumptions, our object semantic description approach follows the flow of conceptual processes presented in Figure 1 as a hypothesis for simulating the human behavior in object features description. Since our proposal requires as *priori knowledge* the prototypes representation of objects categories, we need a procedure to represent the prototype of a specific category.

3.1 Semantic Representation

Category semantic structure (*i.e.*, *central and peripheral meaning*) is related with differences of typicality and membership salience of category members (*extensional non-equality*). The prototype can be understood as the “average” of the abstractions of all objects in the category (Sternberg and Sternberg 2016); it summarizes the most representative members (or features) of the category. The combination between observed object features and features relevance for the category enables the grouping of objects into family resemblance (*intensional non-equality*). This approach justifies the object’s position within the semantic structure of the category and allows typical objects to be grouped into the semantic center of the category (*prototypical organization*).

Let O be an *universe of objects*; $C = \{c_1, c_2, \dots, c_n\}$ be the finite set of objects categories labels that partition O ; $O_{c_i} = \{o \in O : \text{category}(o) = c_i\}$ is the set of objects that share the same i -th category $c_i \in C$, $\forall i = 1, \dots, n$; and $F = \{f_1, f_2, \dots, f_m\}$ be a finite set of distinguishing features of an object.

Definition 1 *Semantic prototype*. We call the *central meaning* of the category $c_i \in C$, *semantic prototype* of c_i -category, or simply *semantic prototype*, to the “average” and standard deviation of each features of all *typical objects* within the c_i -category, along with a measure of the relevance of those features. Formally, our semantic prototype is a 3-tuple $P_i = (M_i, \Sigma_i, \Omega_i)$ where $\forall i = 1, \dots, n; \forall j = 1, \dots, m$:

- i) $M_i = [\mu_{i1}, \mu_{i2}, \dots, \mu_{im}]$ is a nonempty m -dimensional vector, where μ_{ij} is the *mean* of j -th feature of features extracted for *only typical objects* of c_i -category;
- ii) $\Sigma_i = [\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{im}]$ is a nonempty m -dimensional vector, where σ_{ij} is the standard deviation of j -th feature of features extracted for *only typical objects* of c_i -category;
- iii) $\Omega_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{im}]$ is a nonempty m -dimensional vector, where ω_{ij} is the relevance value of j -th feature for the category $c_i \in C$.

Definition 2 *Abstract prototype*. The abstract semantic center of i -th category $c_i \in C$, most prototypical element of i -th category, ideal element of i -th category, or simply the *abstract prototype* of i -th category, is the m -dimensional vector $M_i \in P_i = (M_i, \Sigma_i, \Omega_i)$ composed of the expected value of each features extracted for *only typical objects* of c_i -category.

3.2 Semantic Distance

Our description approach (see processes 4 - 5 in Figure 1) needs a distance measure to compute the discrepancy between object features and category-typical features (semantic prototype). The distance metrics $L1$ and $L2$ could be good options if it did not assume that all object features have the same relevance.

According to the Prototype Theory: *i*) each object feature has a relative relevance in the category and *ii*) the relevance (or salience) of each category member is in accordance with the number and type of features present in the object. This approach can establish a degree of prototypicality of a specific element within the category (*extensional non-equality*).

Some formal models of Experimental Psychology such as *prototype model* (Reed 1972; Homa and Vosburgh 1976), *Multiplicative Prototype Model* (MPM) (Minda and Smith 2001, 2002) and *Generalized Context Model* (GCM) (Medin and Schaffer 1978; Estes 1986; Nosofsky 1986; Zaki et al. 2003) proposed measures of semantic distances between stimulus that correspond to Prototype Theory foundations. Consequently, we generalized some of these semantics measures to propose a *semantic distance metric* (or dissimilarity function) that measures the discrepancy between two objects images (or between an object image and its semantic prototype) based on observed features.

Definition 3 *Distance between objects*. Let $o_1, o_2 \in O_{c_i}$ be a representative objects of i -th category $c_i \in C$; F_{o_1}, F_{o_2} the features of objects o_1, o_2 respectively. We defined the *objects distance* between o_1 and o_2 as the semantic distance given by:

$$\delta(o_1, o_2) = \sum_{j=1}^m |\omega_{ij}| |f_j^1 - f_j^2|, \quad (1)$$

where $\omega_{ij} \in \Omega_i$, $f_j^1 \in F_{o_1}$ and $f_j^2 \in F_{o_2}$, $\forall i = 1 \dots n; \forall j = 1 \dots m$.

It is worth noting that our semantic *distance between objects* is a generalization of the *psychological distance between two stimuli* proposed in GCM formal model. Unlike the original formal *Context Model* (Medin and Schaffer 1978), we assume that: *i*) object features (stimuli) are not binary values ($f_j \in \mathbb{R}$); *ii*) relevance (ω_{ij}) (or cost of attention) of each j -th unitary object feature is forced to be strictly positive, but has no upper limit ($\sum_{j=1}^m \omega_{ij} \neq 1$). We removed these constraints of GCM Model in order to model *object features* and *object features relevance* using the features and weights learned by classification models.

Definition 4 *Prototypical distance.* Let $o \in O_{c_i}$ a representative object of i -th category $c_i \in C$, F_o the features of object o and $P_i = (M_i, \Sigma_i, \Omega_i)$ the semantic prototype of c_i -category. We defined as *prototypical distance* between o and P_i the semantic distance:

$$\delta(o, P_i) = \sum_{j=1}^m |\omega_{ij}| |f_j - \mu_{ij}|, \quad (2)$$

where $\omega_{ij} \in \Omega_i$, $\mu_{ij} \in M_i$, and $f_j \in F_o$; $M_i, \Omega_i \in P_i$ $\forall i = 1..n$; $\forall j = 1..m$.

Our prototypical distance is a generalization of semantic distance of MPM formal model (Minda and Smith 2001, 2002). Different from MPM model assumptions, we assumed that prototype features are not features of a real member of i -th category, but features of expected ideal member (our *abstract prototype*) of i -th category ($M_i \in P_i$).

Definition 5 *Features metric space.* Let F_{c_i} be a non empty set of all objects features of category $c_i \in C$. Since the distance function $\delta : F_{c_i} \times F_{c_i} \rightarrow \mathbb{R}^+$ satisfies the axioms of *non-negativity*, *identity of indiscernible*, *symmetry* and *triangle inequality*; δ is a *metric* in the features domain F_{c_i} . Consequently, (F_{c_i}, δ) is a *metric space* or *features metric space*.

Proof Let $o_1, o_2, o_3 \in O_{c_i}$ objects members of i -th category ($c_i \in C$); F_1, F_2, F_3 the corresponding object features with $f_j^1 \in F_1, f_j^2 \in F_2, f_j^3 \in F_3; \forall i = 1, \dots, n; \forall j = 1, \dots, m$.

- $\delta(o_1, o_2) \geq 0$ (*non-negativity*).
Since all terms in Equation 1 are non negative (≥ 0), $\delta(o_1, o_2) \geq 0$ by definition;
- $\delta(o_1, o_2) = 0 \Leftrightarrow o_1 = o_2$ (*identity of indiscernible*).

– $\delta(o_1, o_2) = 0 \rightarrow o_1 = o_2$.
If $\delta(o_1, o_2) = 0$ then $\sum_{j=1}^m |\omega_{ij}| |f_j^1 - f_j^2| = 0$; consequently, since all terms in Equation 1 are non negative, the above expression is true if each element in the sum is zero. Then, $\forall |\omega_{ij}| \neq 0, |f_j^1 - f_j^2| = 0 \rightarrow f_j^1 = f_j^2 \rightarrow o_1 = o_2$;

– $o_1 = o_2 \rightarrow \delta(o_1, o_2) = 0$.
If $o_1 = o_2 \rightarrow f_j^1 = f_j^2 \rightarrow |f_j^1 - f_j^2| = 0, \forall j = 1..m$; then $\delta(o_1, o_2) = 0$;

- $\delta(o_1, o_2) = \delta(o_2, o_1)$ (*symmetry*).
 $\delta(o_1, o_2) = \sum_{j=1}^m |\omega_{ij}| |f_j^1 - f_j^2| = \sum_{j=1}^m |\omega_{ij}| |f_j^2 - f_j^1| = \delta(o_2, o_1)$;
- $\delta(o_1, o_3) \leq \delta(o_1, o_2) + \delta(o_2, o_3)$ (*triangle inequality*).
 $\delta(o_1, o_2) + \delta(o_2, o_3) = \sum_{j=1}^m |\omega_{ij}| |f_j^1 - f_j^2| + \sum_{j=1}^m |\omega_{ij}| |f_j^2 - f_j^3| = \sum_{j=1}^m |\omega_{ij}| (|f_j^1 - f_j^2| + |f_j^2 - f_j^3|)$

and by absolute value property $|f_j^1 - f_j^2| + |f_j^2 - f_j^3| \geq |f_j^1 - f_j^3|$, then $\delta(o_1, o_2) + \delta(o_2, o_3) \geq \delta(o_1, o_3)$.

Also, note that if $E \subseteq O_{c_i}$ is a subset of i -th category, $\delta(E) = \sum \delta(o, P_i), \forall o \in E$. Consequently, our prototypical distance satisfies the following properties: *i) null empty set*: $\delta(\emptyset) = 0$; *ii) countable additivity*: for all countable collections $\{E_k\}_{k=1}^{\infty}$ of pairwise disjoint sets in E , $\delta\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \delta(E_k)$ (this property is easy to prove using mathematical induction).

Corollary 1 *The prototypical distance function from F_{c_i} to the extended real number line, $\delta : F_{c_i} \rightarrow \mathbb{R}^+$, is a measure. Consequently, (F_{c_i}, δ) is a measurable space.*

Since our statement of (F_{c_i}, δ) is a measurable space, we can use the generalization of Chebyshev's inequality (Chebyshev 1867) to define the boundary of our semantic prototype representation. Chebyshev (1867) asserted that the probability that a scalar random variable ξ with distribution \Pr differs from its mean $\mu \in \mathbb{R}$ by more than $\lambda \in \mathbb{R} > 0$ standard deviations $\sigma \in \mathbb{R} > 0$ satisfies the relation: $\Pr(|\xi - \mu| \geq \lambda\sigma) \leq \min(1, \frac{1}{\lambda^2})$.

Saw et al. (1984) and Stellato et al. (2017) approached the problem of formulating an empirical Chebyshev inequality given N i.i.d samples from an unknown distribution \Pr , and their empirical mean μ_N and empirical standard deviation σ_N . Saw et al. (1984) and Stellato et al. (2017) derives a Chebyshev inequality bound with respect to the $(N + 1)$ -th sample. The Multivariate Chebyshev inequality (Stellato et al. 2017) can define the boundary for an ellipsoidal set centered at the mean.

Definition 6 *Semantic prototype edges.* Let (F_{c_i}, δ) be the metric space of object features of i -th category $c_i \in C$. Let $E \subseteq F_{c_i}$ be a set of features extracted for *only typical objects* of c_i -category, $N = |E|$, and F_o the features of a object $o \in O_{c_i}$. We *weakly* defined as *edges* of our semantic prototype $P_i = (M_i, \Sigma_i, \Omega_i)$, the threshold vector $\vec{\lambda}_i = [\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{im}]$ that meets the expression:

$$\Pr(|f_j - \mu_{ij}| \geq \lambda_{ij} \sigma_{ij}) \leq \min(1, \frac{1}{\lambda_{ij}^2}), \quad (3)$$

where $f_j \in F_o, \mu_{ij} \in M_i$ and $\sigma_{ij} \in \Sigma_i, \forall i = 1..n; \forall j = 1..m$. Finally, given a probability bound, it is possible to compute a threshold vector $\vec{\lambda}_i$ and construct a confidence ellipsoidal set from the sample mean and covariance of only typical objects samples (a stronger *semantic prototype edges* definition can be performed using – completely – the Stellato et al. (2017) statements).

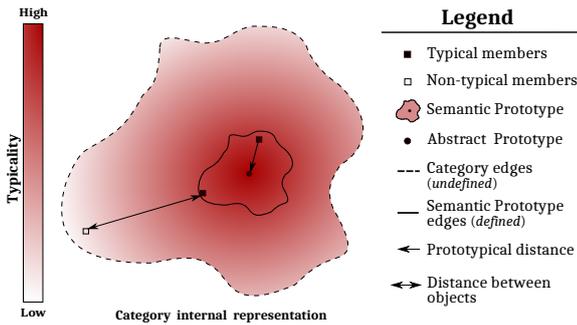


Fig. 3 *Category internal structure.* Figure shows our expected semantic representation of category internal structure. Also we show the principal definitions of our Computational Prototype Model.

Figure 3 shows the expected representation of category internal structure based on our Computational Prototype Model (CPM) [*Semantic prototype* (Definitions 1 and 2) + *Semantic distance* (Definitions 3 and 4)]. With our CPM model, we try to respect some important concepts of the Prototype Theory: *i*) category prototype edges are defined with our vector $\Sigma_i \in P_i = (M_i, \Sigma_i, \Omega_i)$; *ii*) category edges are blurred (not sharp defined) because our semantic prototype is not computed with all category elements (only with typical elements); *iii*) objects representativeness (typicality) within the category is simulated with our prototypical distance.

3.3 Prototype Construction

Our *semantic prototype* representation can be easily computed by any model with the ability to extract object features of images (F_o) and learn the unitary relevance value ($\omega_{i,j}$) of each j -th object feature in i -th category. We have also considered the elements typicality within the category to compute our *semantic prototype*. Consequently, we need objects datasets with annotations of objects typicality scores.

Moreover, our object description approach presented in Figure 1 attempts — following the human behavior — to use the same features extracted to classify and describe objects. First, we need to recognize the category to which the object belongs and then, find what the object features that distinguish it from others within the category are. However, how to model a global object description with similar behavior of the Figure 1 diagram?

To address these issues, we rely on the fact that CNNs provide outstanding performance in image semantic processing and classification tasks. We used CNN-classification models for features extraction, recogni-

Algorithm 1 Prototype Construction

Input: CNN-model Λ , objects dataset O , category c_i
Output: Category Prototype (P_i)
 $O_{c_i} \leftarrow \{o \in O : \text{category}(o) = c_i\}$
 $\text{features_block} \leftarrow \{\}$
for $o \in O_{c_i}$ **do**
 if o *is_typical* **then**
 $F_o \leftarrow \Lambda.\text{features_of}(o)$
 $\text{features_block} \leftarrow \text{features_block} \cup F_o$
 $\Omega_i, b_i \leftarrow \Lambda.\text{softmax_weight_learned_of}(c_i)$
 $M_i, \Sigma_i \leftarrow \text{compute_stats}(\text{features_block})$
return ($M_i, \Sigma_i, \Omega_i, b_i$)

tion, and classification of the visual information received as input (processes 1 to 4 in Figure 1). CNN-models, analogous to the human memory (Fuster 1997), make associations that keep the knowledge in its connection structures. Our method downloads that knowledge of pre-trained CNN-models into a semantic structure (*semantic prototype*), which aims is to stand for the central semantic meaning of learned categories (see step 5 in Figure 1).

Definition 7 *Convolutional semantic prototype.* The *convolutional semantic prototype* of i -th category $c_i \in C$ is a 4-tuple $P_i = (M_i, \Sigma_i, \Omega_i, b_i)$, where M_i, Σ_i are computed using features of c_i -category extracted from the *fully convolutional layer* of pre-trained CNN - classification models; and Ω_i, b_i are the learned parameters (learned features relevance) of i -th category in the softmax layer. Next, we refer to *convolutional semantic prototype* of the category as a *semantic prototype*.

Algorithm 1 details the computation of a *semantic prototype* for a specific category. Given a labeled object images dataset, for each object category in dataset, we use Algorithm 1 to compute the correspond semantic prototype (*off-line processing*). The resulting *semantic prototypes dataset* is used as *prior knowledge* in our prototype-based description model (see Figure 1). Figure 4 shows the main steps and concepts of our prototype construction algorithm.

Semantic prototype visualization

Visual representation of semantic prototypes allows presenting a visual summary of category typical features. Some approaches (Wohllhart et al. 2013; Li et al. 2018) learn prototypes representations in the image domain; and consequently, prototype visualization is simply the image learned. These approaches require a considerable computational expense to learn its prototypes visualization. Wohllhart et al. (2013) introduced the learning of image prototypes representations in the back-propagation process. On the flip side, Li et al. (2018)

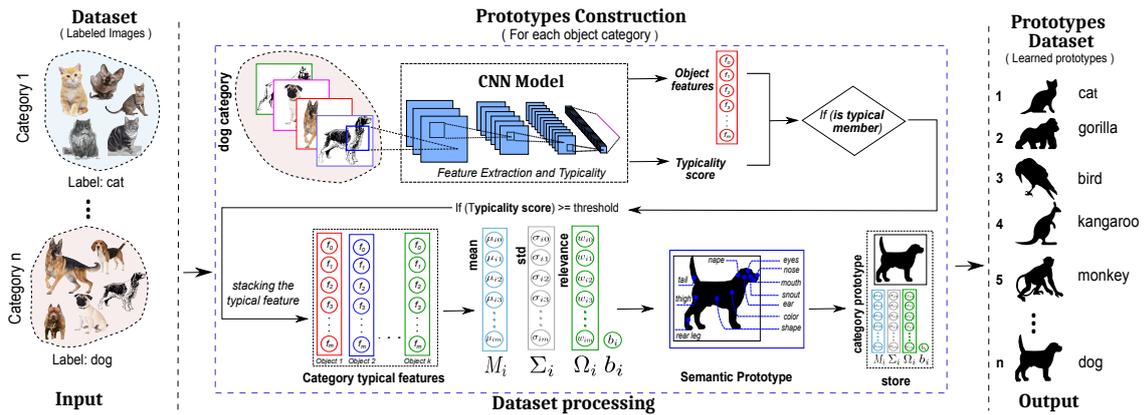


Fig. 4 Off-line construction of the semantic prototypes dataset. Given a labeled images dataset, for each objects category present in the dataset, we compute our semantic prototype representation using Algorithm 1.

used an encoder-decoder deep architecture to learn prototypes visualization. Since our semantic prototype representation is constructed straightforwardly from pre-trained CNN classification models, the methods mentioned above are not appropriate to visualize our prototype representation.

Binder et al. (2016) proposed a circular visualization of semantic mean attribute vectors for concrete object noun categories. Consequently, a simple approach for visualizing our semantic prototype representation is to visualize the distribution values of each m -dimensional vector that compose the P_i -tuple of our semantic prototype definition.

Figure 5 shows an illustration of our semantic prototype representation corresponding to i -th category. We showed each tuple-member (m -dimensional vector) that composes the proposed semantic prototype of i -

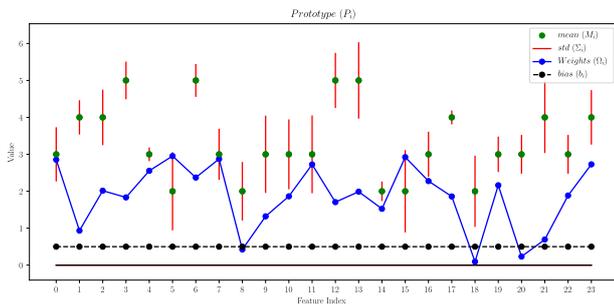


Fig. 5 Visualization of our semantic prototype representation $P_i = (M_i, \Sigma_i, \Omega_i, b_i)$ of i -th category. We showed the m -dimensional vector M_i (mean of *typical* members features) and m -dimensional vector Ω_i (measure of features relevance within i -th category) in *green* and *blue* colors, respectively. The m -dimensional vector Σ_i (standard deviation of *typical objects* features of i -th category) is represented as feature boundary (in red lines) for each j -th unitary feature. Learned bias value b_i is represented as a m -dimensional vector.

th category, $P_i = (M_i, \Sigma_i, \Omega_i, b_i)$. We represented the learned bias value b_i as the bias m -dimensional vector $\vec{b}_i \in \mathbb{R}^m$, $\vec{b}_i = \frac{b_i}{m} \cdot \vec{1}$ such that $b_i = \sum_m \vec{b}_i$.

It is noteworthy that our semantic prototype has a values distribution that is characteristic of i -th category it represents. *I.e.*, our *semantic prototype* can be understood as a DNA chain that stands for the category members' typical features. The semantic prototype representation uniqueness is guaranteed by the relevance vector (Ω_i), which was learned specifically for that i -th category when the CNN-classification model was trained.

4 Global Semantic Descriptor

In the previous section, we presented a framework to encapsulate the central meaning (semantic prototype) of an object category. In this section, we present how to introduce that semantic prototype representation to simulate the object semantic description work-flow depicted in Figure 1.

4.1 Semantic Meaning

Some cognitive neuroscience researches have studied the effect of *semantic meaning* in object recognition task (Tulving 2007; Martin 2007; Collins and Curby 2013). When an object has been previously associated with some type of semantic meaning in the brain, people are more prone to identify the object (Tulving 2007; Martin 2007) correctly. Studies (Tulving 2007; Martin 2007; Collins and Curby 2013) have shown that semantic associations allow a much faster recognition of an object, even when the task of object recognition becomes increasingly difficult (varying points of view, oc-

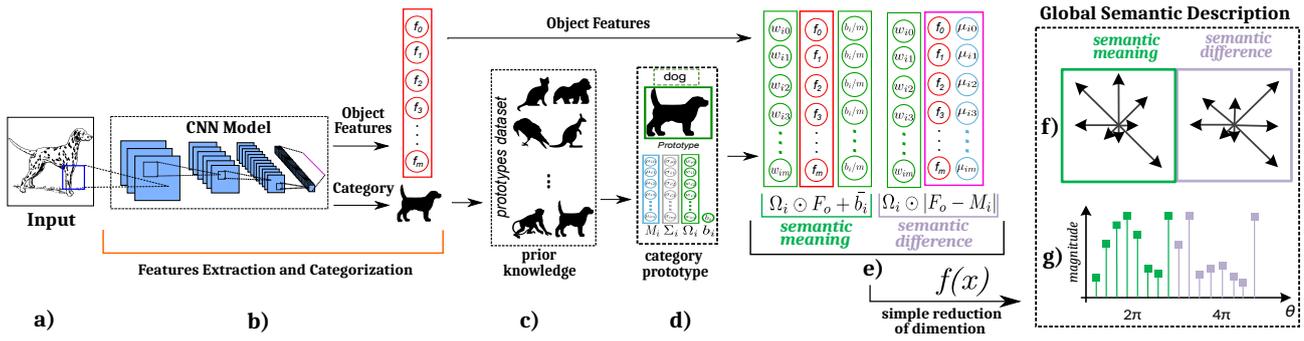


Fig. 6 Overview of our prototype-based description model. Set of steps to transform the visual information received as input into a Global Semantic Descriptor signature. a) input image; b) features extraction and classification using a CNN-classification model; c) prototypes dataset; d) category prototype selection; e) global semantic description of object using category prototype; f) graphic representation of our Global Semantic Descriptor signature resulting from the dimensionality reduction function ($f(x)$); and g) Global Semantic Descriptor signature.

clusion) (Collins and Curby 2013). Therefore, semantic associations based on object semantic meaning allow for faster object recognition.

Moreover, the fact that some CNN models (e.g., ResNet (He et al. 2016)) outperform the human-reported performance (5.1% (Russakovsky et al. 2015)) on large-scale visual object classification tasks, generated some cognitive studies (Yamins et al. 2014; Cadieu et al. 2014; Khaligh-Razavi and Kriegeskorte 2014; Cichy et al. 2017) to research the possible links between CNN models and visual system in the human brain. Cichy et al. (2017) suggested that deep neural networks perform spatial arrangement representations like those performed by a human being. Khaligh-Razavi and Kriegeskorte (2014) concluded that the weighted combination of features in the last fully connected layer of CNN models could thoroughly explain the inferior temporal cortex in the human brain. We lay hold of these theoretical foundations to model our representation of *objects semantic meaning*.

Definition 8 *Semantic value*. Let be F_o observed features of an object $o \in O$ ($F_o = \{f_1, f_2, \dots, f_m\}$). The *semantic meaning* of object features F_o for category $c_i \in C$, *summary value* of features F_o , or simply *semantic value* of F_o in c_i -category is an abstract value: $z = \sum_m \omega_{ij} f_j + b_i$, where $\omega_{ij} \in \Omega_i$, $f_j \in F_o$. Consequently, the semantic value of ideal member of c_i -category, *central semantic meaning* of c_i -category or *summary value* of the semantic prototype $P_i = (M_i, \Sigma_i, \Omega_i, b_i)$ is the *semantic value* $\hat{z}_i = \sum_m \omega_{ij} \mu_j + b_i$, where $\omega_{ij} \in \Omega_i$, and $\mu_{ij} \in M_i$ are the *abstract prototype* features, $\forall i = 1, \dots, n$; $\forall j = 1, \dots, m$.

Note that our *object semantic value* is exactly the same value used to object categorization in softmax layer of CNN-classification models. Hence, our approach

of object semantic description based on prototypes assumes as object *semantic meaning* vector, the semantic vector ($\vec{z} = \Omega_i \odot F_o + \vec{b}_i$) constructed with the element-wise operations to compute the object *semantic value* (Definition 8). Our *semantic meaning* representation uses a bias vector (\vec{b}_i) to uniformly dissolve the bias value in each semantic vector component ($b_i = \sum_m \vec{b}_i$). With this approach it is enough a sum of each *semantic meaning vector* component to recover the *object semantic value* ($z = \sum_m \vec{z}$). Accordingly, our *semantic meaning vector* contains the same semantic definitions used for CNN models to categorize an object within a specific category.

4.2 Semantic Difference

We stand for the *semantic distinctiveness* of an object for specific c_i -category as the semantic discrepancy between object features and features of the most prototypical (ideal) element of c_i -category (abstract prototype of c_i -category). Since object features (F_o) and *abstract prototype* of c_i -category ($M_i \in P_i$) belong to the same features domain (features metric space), we apply our *prototypical distance* as measure of the objects distinctiveness within a category.

Consequently, our approach assumes as object *semantic distinctiveness vector*, the *semantic difference vector* ($\vec{\delta} = \Omega_i \odot |F_o - M_i|$) constructed with the element-wise operations to compute the object *prototypical distance* (Definition 4). Our *semantic difference vector* is the weighted (Ω_i) *residual vector* ($\vec{r} = |F_o - M_i|$) composed of absolute values of the difference between each object feature and each feature of c_i -category abstract prototype.

Note that our *object semantic difference* (or our prototypical distance) can be understood as the sum of

Algorithm 2 Global Semantic Descriptor ψ

-
- 1: **Input:** Image of an object o
 - 2: **Output:** Object semantic signature (ψ_o)
 - 3: **Prior Data:** Trained CNN-model A , $prototypes_dataset$
 - 4: $F_o, c_i \leftarrow A.features_and_prediction(o)$
 - 5: $M_i, \Sigma_i, \Omega_i, b_i \leftarrow prototypes_dataset(c_i)$
 - 6: $meaning \leftarrow f(F_o, \Omega_i, b_i, meaning)$
 - 7: $difference \leftarrow f(|F_o - M_i|, \Omega_i, b_i, distinctiveness)$
 - 8: **return** $meaning \oplus difference$
-

absolute difference between the *object semantic meaning vector* (\vec{z}) and the *central semantic meaning vector* (\vec{z}_i) of c_i -category. Thus, Equation 2 is equivalent to $\sum_{j=1}^m |\vec{z}_j - \vec{z}_{ij}| = \sum_{j=1}^m |\omega_{ij} f_j - \omega_{ij} \mu_{ij}| = \delta(o, P_i)$ when $\forall \omega_{ij} \in \Omega_i, \omega_{ij} \geq 0$ (we introduced this ω_{ij} constraint in the semantic distance of MPM model). Therefore, our *object semantic difference* representation has the advantage that elements vector sum is enough to retrieve the object *prototypical distance* ($\delta = \sum_m \vec{\delta}$).

Figure 6 depicts an overview of our prototype-based description model. Our *Global Semantic Descriptor based on Prototypes (GSDP)* uses as a requirement the *prior knowledge* of each category prototype (prototypes are precomputed off-line using Algorithm 1). After feature extraction and categorization processes (Figure 6b), we use the corresponding category prototype for semantic description of object features. We show in Figure 6e) the steps to introduce the category prototype into the global semantic description of object’s features. A drawback of our object semantic representation (Figure 6e) is having high dimensionality, since it is based on *semantic meaning vector* (\vec{z}) and *semantic difference vector* ($\vec{\delta} = \Omega_i \odot \vec{r}$). The large dimensionality of our feature vectors might make its use unfeasible in common computer vision tasks (Han et al. 2017; Kim et al. 2018). Figure 6 and Algorithm 2 detail the main steps of our approach; note that steps follow the same work-flow of human description hypotheses depicted in Figure 1.

4.3 Dimensionality Reduction

Several dimensionality reduction algorithms such as PCA (Abdi and Williams 2010) and NMF (Lee and Seung 2001) are based on discarding features that do not generate a meaningful variation. Although these approaches work on some tasks, after applying these algorithms, we lost the ability of data interpretation (Abdi and Williams 2010). From the Prototypes Theory perspective, discarding features is no suitable when it is applied to the semantic space due to the absence of necessary and sufficient definitions to categorize an object (*intentional non-discreteness*). Occasionally when discarding features might lead in discarding elements of the cate-

gory (Geeraerts 2010). For instance, there may be some objects within the category that do not have some category typical features (flying is a typical feature of bird category; however, a penguin is a bird that does not fly).

We proposed a simple transformation function $f(x)$ to compress our global semantic representation of the object’s features (Figure 6e) in a low dimensional global semantic signature (Figure 6g). Our transformation function aims to reduce our semantic representation dimensionality while keeping the property of *easy retrieve* the *object semantic meaning* and *object semantic difference* from the final descriptor signature. Our final descriptor signature (ψ) is computed by concatenating the corresponding signatures of *semantic meaning vector* (\vec{z}) and *semantic difference vector* ($\vec{\delta}$) compressed with our $f(x)$ transformation (see Algorithm 2).

Figure 7 shows the main steps of our $f(x)$ transformation. We use a square auxiliary matrix ($\chi_{r \times r}$) as a parameter to control the descriptor signature dimensionality. The auxiliary matrix dimensions is a parameter that allows us to control the final GSDP-signature size, *i.e.*; larger auxiliary matrix dimensionality leads to smaller GSDP signature; and vice versa.

The main steps showed in Figure 7 can be summarized as: **1)** Resize the input vectors in the best 2D dimensional configuration of matrices ($p \times q$) whose dimensions are multiples of r (auxiliary matrix dimension). **2-3)** Compute the angles matrix ($\Theta_{r \times r}$) with angles formed by the position of each feature with respect to auxiliary matrix $\chi_{r \times r}$ center; to achieve uniqueness the diagonal angles were evenly distributed between the magnitudes of angles α and β . **4) - 5)** Create *unitary semantic gradient* for each auxiliary matrix mapped within $p \times q$ matrices; each *semantic gradient* is constructed using the *angle matrix* ($\Theta_{r \times r}$), and *magnitude* and *sign* of semantic vectors computed using Definition 2 and 8. **6)** Reduce the semantic gradient to 8-vectors similarly to SIFT approach (Lowe 2004); **7)** Concatenate, for each auxiliary matrix $\chi_{r \times r}$ mapped, the corresponding unitary 8D-signatures resulted of flow 4-6. Algorithm 3 details all steps.

Hence, our final descriptor signature preserves the *object semantic meaning* (Property 1) and the *object semantic difference* (Property 2) presented in our first global semantic representation of object features (Figure 6e). Additionally, depending of the input vector, our descriptor can uses $f(x)$ transformation to construct global semantic representations (signatures) with different meanings within i -th category (Property 3). In other words, our descriptor can construct semantic representations (see Figure 6e) for: *i)* an object, *ii)* ideal category member (abstract prototype), and *iii)* cate-

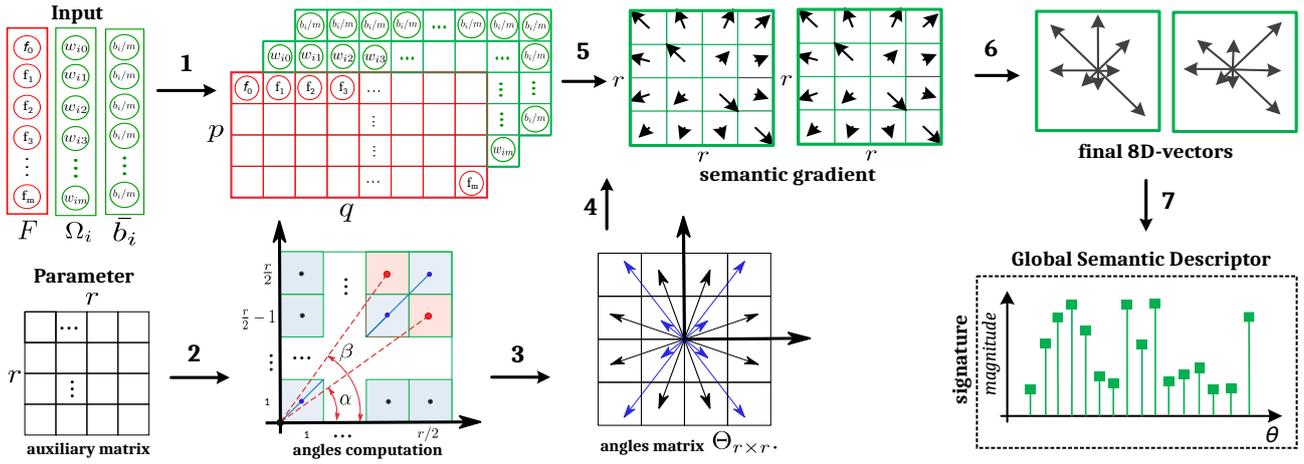


Fig. 7 Dimensionality reduction function. Figure shows our transformation $f(x)$ to convert the high dimensionality of our object semantic representation into the corresponding semantic descriptor signature. Final signature is constructed by concatenating each 8D-vector computed from each unitary semantic gradient. We showed the trivial case when the input m -dimensional vectors have 2 times auxiliary matrix dimension ($m = p \cdot q$ and $p = r$; $q = 2r$); consequently, output signature has 2 times (16D) the 8D-vector dimension.

Algorithm 3 Dimensionality Reduction $f(x)$

- 1: **Input:** m -dimensional vector α , Ω_i, b_i , type
 - 2: **Output:** Semantic signature
 - 3: **Parameter:** Auxiliary matrix $\chi_{r \times r}$
 - 4: $\bar{b}_i \leftarrow \frac{b_i}{m}$ // m -dimensional vector \bar{b}_i ($b_i = \sum_m \bar{b}_i$)
 - 5: $\chi_{r \times r} \leftarrow \text{shape}(r, r)$ // setting auxiliary matrix dimension
 - 6: Computing angles matrix: $\Theta_{r \times r} = \text{angles_from}(\chi_{r \times r})$
Find new shape p, q from k
 - 7: Finding the optimal configuration p, q where $p \equiv 0 \pmod{r}$, $q \equiv 0 \pmod{r}$ and $m = p \cdot q$
 - 8: $\alpha, \Omega_i, \bar{b}_i = \text{reshape_to_matrix}_{p \times q}(\alpha, \Omega_i, \bar{b}_i)$
 - 9: $\text{signature} \leftarrow []$
 - 10: **for** $j = 1, \dots, \frac{q}{r}; k = 1, \dots, \frac{p}{r}$ **do**
 - 11: Mapping $\chi_{r \times r}$ in $\alpha, \Omega_i, \bar{b}_i$
 - 12: Computing \vec{z}_i^{jk} using Hadamard product \odot .
 - 13: $\vec{z}_i^{jk} = \begin{cases} \Omega_i^{jk} \odot \alpha^{jk} + \bar{b}_i^{jk}, & \text{if type} = \text{meaning} \\ |\Omega_i^{jk}| \odot \alpha^{jk}, & \text{otherwise} \end{cases}$
 - 14: $g^{jk} \leftarrow \text{vectors}(\Theta_{r \times r}, |\vec{z}_i^{jk}|, \text{sign}(\vec{z}_i^{jk}))$.
 - 15: $\text{signature}^{jk}(l) = \sum g^{jk}(\theta), \forall \theta \in \Theta_{r \times r} : \theta_l - 45 < \theta \leq \theta_l$ with $\theta_l = l \cdot \frac{\pi}{4}, \forall l = 1, \dots, 8$
 - 16: $\text{signature} \leftarrow \text{signature} \oplus \text{signature}^{jk}$
 - 17: **return** signature
-

gory semantic meaning encapsulated with semantic prototype boundaries.

4.4 Descriptor Properties

Property 1 Semantic meaning preservation. The semantic descriptor signature preserves the *object semantic value*: $\sum_{l=0}^{|\psi|/2} \psi[l] = \hat{z}$.

Proof To prove this, it suffices to follow backward through steps 6 and [9, 17] of Algorithm 2 and 3, respectively. $\sum_{l=0}^{|\psi|/2} \psi = \sum f(\alpha, \Omega_i, b_i, \text{meaning}) = \sum \sum_j \sum_k g^{jk} = \sum \Omega_i \odot \alpha + \bar{b}_i = \sum \vec{z} = \hat{z}; \alpha \in \{M_i, F_o\}$.

Property 2 Prototypical distance preservation. If $o \in O_{c_i}$ is a object of i -th category, the object signature ψ_o preserves the object *prototypical distance*: $\sum_{l=|\psi_o|/2}^{|\psi_o|} \psi_o[l] = \delta(o, P_i)$.

Proof Similar to the previous proof, but using distinctiveness vector ($|\Omega_i| \odot |F_o - M_i|$) through steps 7 and [9, 17] of Algorithm 2 and 3, respectively. $\sum_{l=|\psi|/2}^{|\psi|} \psi = \sum f(|F_o - M_i|, \Omega_i, b_i, \text{distinctiveness}) = \sum \sum_j \sum_k g^{jk} = \sum |\Omega_i| \odot |F_o - M_i| = \sum \vec{\delta} = \delta(o, P_i)$.

Property 3 Structural polymorphism. Our global semantic descriptor GSDP has the polymorphic property of describing, with the same structural representation, distinctly different semantic meanings within the c_i -category. Thus, our descriptor uses the category prototype $P_i = (M_i, \Sigma_i, \Omega_i, b_i)$ to construct different semantic signature taxonomies:

- i) an object $o \in O_{c_i}$, $\psi_o = \psi(F_o, |F_o - M_i|, \Omega_i, b_i) = f(F_o, \Omega_i, b_i, \text{meaning}) \oplus f(|F_o - M_i|, \Omega_i, b_i, \text{distinctiveness});$
- ii) *central semantic meaning* of i -th category (abstract prototype), $\psi_{P_i} = \psi(M_i, |M_i - M_i|, \Omega_i, b_i) = \psi(M_i, \vec{0}, \Omega_i, b_i);$
- iii) *semantic meaning* of i -th category (semantic prototype), $\psi_i = \psi(M_i, \Sigma_i, \Omega_i, b_i)$.

5 Experimental Evaluation

5.1 Experimental Setup

Aside from performing experiments using benchmark image datasets with fixed-size, size-normalized and centered images like MNIST (Lecun et al. 1998) and CIFAR (Krizhevsky and Hinton 2010), we also evaluated our approach on ImageNet (Russakovsky et al. 2015) as real images dataset. For each image dataset, we used a CNN-classification model for feature extraction and classification (see Figure 6b). Thus, we used a CNN-MNIST and CNN-CIFAR models based on *LeNet* (Lecun et al. 1998) and *Deep Belief Network* (Krizhevsky and Hinton 2010) architectures for image classification in MNIST and CIFAR datasets, respectively. Also, we conducted experiments in ImageNet using VGG16 (Simonyan and Zisserman 2014) and ResNet50 (He et al. 2016) models as background of our global semantic description model. Note that our *prototype-based description model* depicted in Figure 6, is scalable and can easily be adapted to any other CNN-classification model.

Prototypes Dataset Construction

Our *prototype-based description model* requires *prototypes dataset* as category *prior knowledge* (see Figure 6c) to build object semantic representations (see Figure 6e) that stand for the object distinctiveness within the category. In the experiments, we computed prototypes datasets with CNN-MNIST, CNN-CIFAR, VGG16, and ResNet50 models in MNIST, CIFAR, and ImageNet datasets, respectively.

For feature extraction, we assumed as object features those extracted from the last dense layer (before the softmax layer) of the CNN-model. Our approach needs typical objects of categories or any information about typicality score (or typicality degree) of objects belonging a specific category to build the proposed semantic prototype properly. However, none of the images datasets used have this annotation. Lake et al. (2015) showed that the output of the last layer of CNN models could be used as a signal for how typical is an input image. Consequently, we used as *typicality score* of objects the strength of classification response to the category of interest. Specifically, we assumed as typical members of a category those elements that are — unequivocally — classified as category members (*typicality score* > 0.99) by CNN models (see Figure 4). Finally, for each category in datasets, we extracted features of typical members and computed the correspond *semantic prototype* (see Definition 1) using Algorithm 1.

5.2 The Semantics behind our Computational Prototype Model

Achieving the member’s prototypical behavior within a category is one of the motivations and theoretical basis of our approach. Nevertheless, there is no defined metric to quantify whether our representation correctly captures the category semantic meaning. This lack of a metric is a consequence of the fact that there is no defined metric to evaluate the object typicality level within a category robustly; this skill is still reserved only for human beings.

In this section, we analyzed the semantics captured by our CPM Model (*semantic prototype + prototypical distance*). The CPM model pursues two main goals: *i*) capture, with the semantic prototype, the central semantic meaning of a specific object category; *ii*) simulate, in a comparable way to human beings, that visually typical elements of category are organized close (based on our prototypical distance metric) to category prototype. Since we do not have annotated images with the object typicality score to robustly evaluate the semantic captured by our representation, we used another approach to analyze the semantics behind our CPM model.

5.2.1 Central and Peripheral meaning

In this section, we analyzed the *central and peripheral* meaning captured by our CPM model. Since we defined *abstract prototype* as the *abstract semantic center* of category, we observed how relevant (or visually representative) – for the category – are those elements allocated by our CPM model in *center and periphery* of category. Expressly, this experiment aims to know what is the visual representativeness of category members closest and furthest from the category semantic center (our abstract prototype).

To achieve such a goal, we extracted object image features using a CNN model and computed our *prototypical distance* for all members of *i*-th category. Finally, the objects images are sorted in ascending order based on the prototypical distance value of each element. Figures 8 and 9 present some examples of central and peripheral meaning captured by our CPM model for images categories of ImageNet and MNIST datasets using VGG16 and CNN-MNIST models – respectively – as image feature extractors.

Notice that our proposal for object image semantic interpretation – using our CPM model in CNN image feature domain – attempts to assign a visual representativeness value (or typicality) to object image within the category to which it belongs.

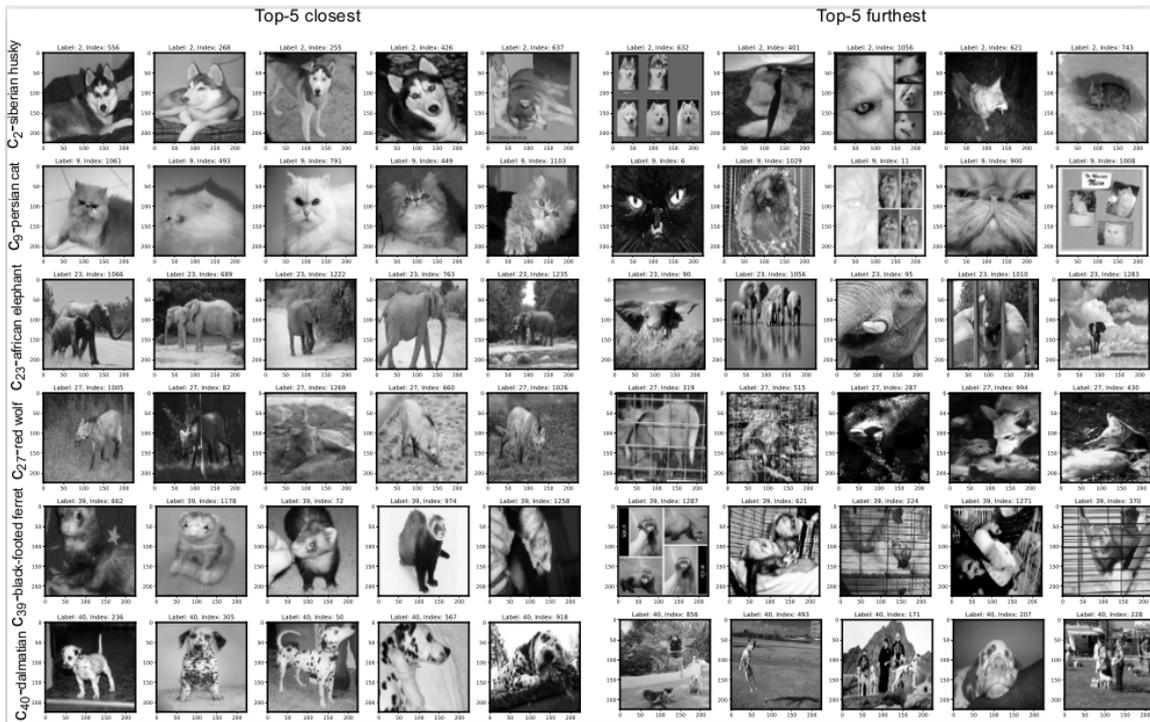


Fig. 8 (from left to right) Top-5 most relevant members identified by our CPM model for a categories sample of ImageNet dataset. (left) Top-5 elements closest to semantic prototype of corresponding category; index value represents the element position within the category dataset. (right) Top-5 elements furthest from the semantic prototype of the category. Object image features was extracted with VGG16 model.

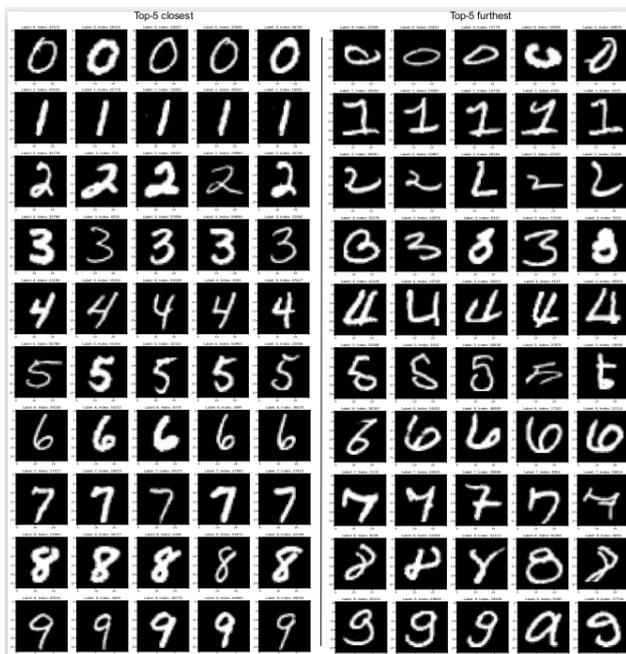


Fig. 9 (from left to right) Top-5 most relevant members identified by our CPM model for all MNIST dataset categories. (left) Top-5 elements closest to category abstract prototype of corresponding category; (right) Top-5 elements furthest from the category semantic prototype. Image features was extracted with CNN-MNIST model.

Figure 9 shows the *Top-5 closest* and *Top-5 furthest* elements from category center (abstract prototype) detected by our CPM model in MNIST categories. For instance, our proposal finds as *typical* elements (Top-5 closest) of *number three* category the handwritten digits with features that are, undoubtedly, distinctive of c_3 -category. Our CPM model also can find the peripheral meaning of the category. Members with fewer characteristic features of *number three*, or little readable, are placed in the periphery (Top-5 furthest) away from the *central semantic meaning*, but keeping the category features (it still belongs to the category). Similar to a human being, our CPM model can find the Top-5 furthest members of *number three* category that are a number 3, but not a typical number 3.

Figure 8 presents the semantic interpretation of visual image information performed by our CPM model in real object images of ImageNet dataset. Note that category members recognized by our CPM model as Top-5 closest members (left column) to category semantic center are easy recognized by human beings, as it exhibits the typical features of an object category. Also, we observed that Top-5 furthest elements (right column) from the semantic prototype (or less representative members of i -th category) detected by

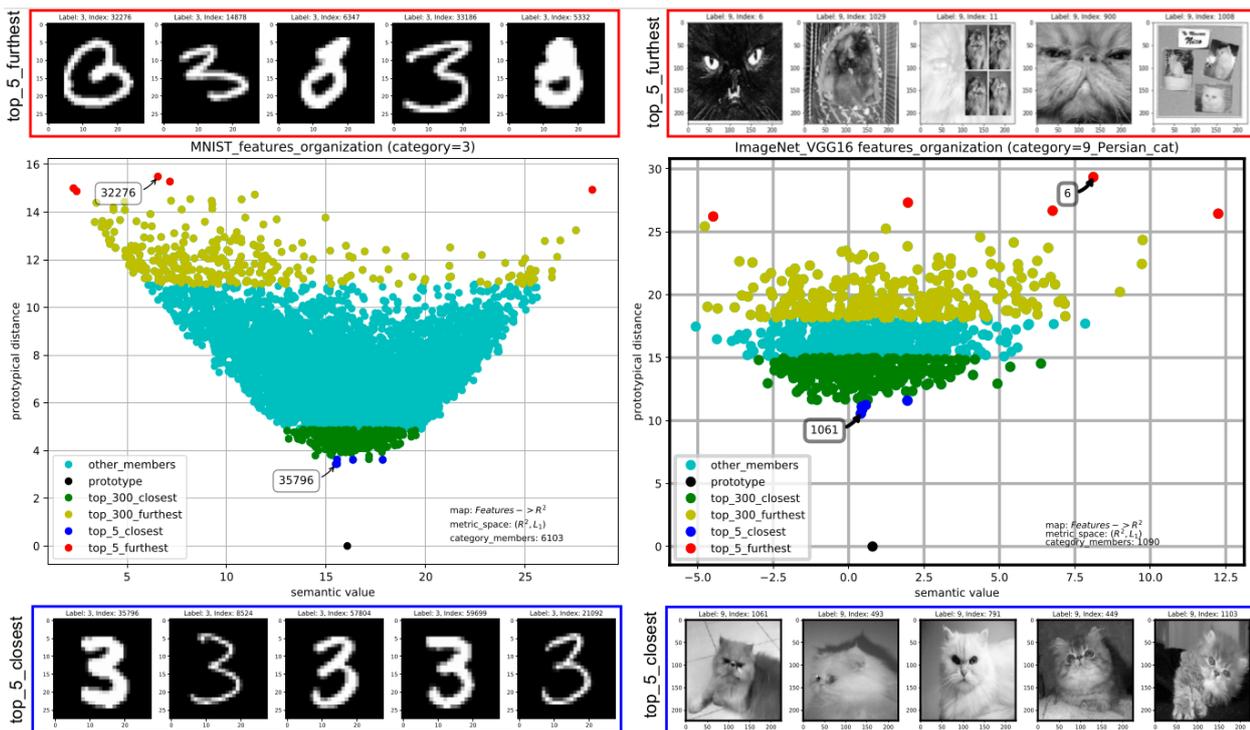


Fig. 10 *Prototypical organization within categories.* Figure shows the internal structure of *number three* and *Persian cat* categories of MNIST and ImageNet datasets, respectively. We represented each category member using image features extracted with CNN-MNIST and VGG16 models. We represented with color degrees the category internal disposition respect its prototype. In bottom and top, from left to right, the mapped Top-5 elements closest (in blue) and furthest (in red) to the mapped semantic prototype (in black) of each category. Image dataset index of the first Top-5 element is annotated inside the black box.

CPM model although retaining some category features are not easily recognized by human beings. That is, our CPM model can identify the most/least visually representative category members, and — correctly — recognizes as category periphery members those elements, where not all typical category features are identified: category typical colors, size, shape, etc. are not easily distinguishable; or object pose in the image does not exhibit these representative features of i -th category. The experiments performed allow to assume that our CPM model can capture the central/peripheral semantic meaning of images categories. But, we still need to answer to the question: Can our CPM model organize all category elements prototypically?

5.2.2 Prototypical Organization

The experiments in this section aim to visualize the internal semantic structure of the category using the semantic meaning encapsulated by our CPM model for each category member. Based on features extracted from objects images, we analyzed the object prototypical behavior observing where it is positioned within the category by our CPM model (using our *prototypical dis-*

tance). Visualizing the semantic position of each category member with respect to *category abstract prototype* constitutes a simple approach to see the internal semantic structure of the entire category structure. We need to corroborate that our CPM model can correctly interpret the object image features and position it semantically within the category, keeping a *prototypical organization* of the category.

Note that our CPM model uses m -dimensional object features from CNN image-features domain. Accordingly, visualizing the category’s internal structure is infeasible in m -dimensional features space since most techniques of data visualization are based on features discarding. From the perspective of the Prototype Theory foundations, features discarding approach can be problematic (*intensional non-discreteness*). For this reason, we used topology techniques to make object image interpretation based on all observed features. We constructed a map function to show that our CPM model can simulate the prototypical organization of members within a category.

Let (F_c, δ) and (\mathbb{R}^2, L_1) be *metric spaces* (see Definition 5), and ρ a function that maps image object features to (\mathbb{R}^2, l_1) metric space using its *semantic value*

and its *prototypical distance*. I.e., $\rho : F_{c_i} \rightarrow \mathbb{R}^2 \mid \rho(o \in O_{c_i}) = \rho(F_o) = p(z_o, \delta(o, P_i))$, where F_o are the object features, z_o is the *object semantic value*, $\delta(o, P_i)$ is the *object prototypical distance*; the point $p(x, y) \in \mathbb{R}^2$ and L_1 is L1-norm condition.

Let be the objects $o_1, o_2 \in O_{c_i}$, and $p_1 = \rho(o_1), p_2 = \rho(o_2)$ be the corresponding mapped points in (\mathbb{R}^2, L_1) metric space. Then, the Sum of Absolute Difference (SAD) between p_1 and p_2 is $L_1(p_1, p_2) = L_1(\rho(o_1), \rho(o_2)) = L_1(p(z_1, \delta(o_1, P_i)), p(z_2, \delta(o_2, P_i))) = |z_1 - z_2| + |\delta_1 - \delta_2|$; using Definitions 3, 4, and 8 we have: $\delta(o_1, o_2) \leq L_1(p_1, p_2) \leq 2\delta(o_1, o_2)$. Consequently, for every $F_{o_1}, F_{o_2} \in F_{c_i}$ and $\varepsilon > 0$, exists a $\varphi = \frac{\varepsilon+1}{2} > 0$ such that: $\delta(o_1, o_2) < \varphi \Rightarrow L_1(\rho(o_1), \rho(o_2)) < \varepsilon$, i.e., ρ is *continuous*. This means that every element of $\rho(o_1)$ neighborhood in (\mathbb{R}^2, L_1) metric space, also belongs into o_1 neighborhood in (F_{c_i}, δ) metric space (if $\rho(o_1) = p_1, \forall p \in \{p_1 \text{ neighborhood}\}, \rho^{-1}(p) \in \{o_1 \text{ neighborhood}\}$). Consequently, the observed behavior of i -th category internal structure – in terms of distance metrics – in (\mathbb{R}^2, L_1) metric space is equivalent to the behavior in feature metric space (F_{c_i}, δ) .

Figure 10 shows an example of the internal semantic structure of MNIST and ImageNet images categories mapped using ρ . Note how Top-5 closest members (based on our *prototypical distance*) are mapped (*in blue*) and positioned near (based on L1 distance) to the mapped abstract prototype (*in black*). The Top-5 most visually representative members of each category in (F_{c_i}, δ) metric space are the same Top-5 most representative (closest to mapped *abstract prototype*) in (\mathbb{R}^2, L_1) metric space. Likewise, the Top-5 fewer representative members (*in red*) continue to be positioned in the category peripheries, far away from the category abstract prototype (our central semantic meaning representation). The experiments show a prototypical organization of mapped members within the category in (\mathbb{R}^2, L_1) metric space. Consequently, based on ρ properties, a similar grouping of objects based on family resemblance is preserved in CNN-features metric space.

Our approach to visualize the category internal structure also allows observing other semantic phenomena related to the object image visual representativeness. The experiments showed that *object semantic value* and *object prototypical distance* place the object image in a unique semantic position within the category internal structure. This result shows that our approach of constructing a semantic object representation (see Figure 6e) based on vector versions of *semantic value* and *prototypical distance* can be able to describe the object image semantically.

5.2.3 Image Typicality Score

We observed that the shape of category internal structure – in (\mathbb{R}^2, L_1) metric space using our visualization approach – strongly depends on semantic values distribution and prototypical distance distribution. Consequently, in this section we analyzed the relationship between *semantic value* and *prototypical distance* variables. Also, we examined how the variations of these variables can influence on object image visual representativeness (typicality) within the category.

Our *prototypical distance* can be understood as the semantic difference between the object semantic meaning and the semantic meaning of category abstract-prototype (see Subsection 4.2). Specifically, if features relevance (Ω_i) of i -th category is strictly positive ($\omega_{ij} \geq 0, \forall \omega_{ij} \in \Omega_i$) then, the variables *prototypical distance* and *semantic value* are – by construction – strongly correlated. However, experiments in MNIST, CIFAR and ImageNet datasets with each corresponding CNN-model showed that there is a small strength of a linear association between those two variables (Pearson coefficient values between -0.3 and 0.3), but it does not conclude that we can generalize a behavioral pattern between *object semantic value* and *prototypical distance*. This Pearson correlation result is consequence of the fact that the weights learned in softmax layer (our feature relevance Ω_i) of CNN models used for feature extraction (CNN-MNIST, CNN-CIFAR, VGG16 and ResNet50) are not strictly positive. Consequently, the *semantic value* is not a strong measure because the addends in the equation can cancel each other out (see Definition 8), and elements with same semantic value does not imply that elements are equal ($z_{o_1} = z_{o_1} \not\Rightarrow o_1 = o_2$).

Lake et al. (2015) showed that *semantic value* can be used as a signal for how typical an input image looks like. In contrast to Lake et al. results, our experiments with VGG16 and ResNet50 models in ImageNet dataset showed that using the *semantic value* as object typicality score can be problematic because objects with same semantic value do not imply same image visual typicality. Figure 10 shows an example of this phenomenon. In *Persian cat* ImageNet category, the 5th element of Top-5 closest to category prototype (in blue) has a semantic value like 2nd position element of Top-5 furthest (in red) (both images *semantic values* are ≈ 2), but objects images are visually different. That is, the *semantic value* could be a necessary condition to image typicality representation, but it is not enough. On the other hand, note how our *prototypical distance* can capture the visual typicality difference between those two objects images.

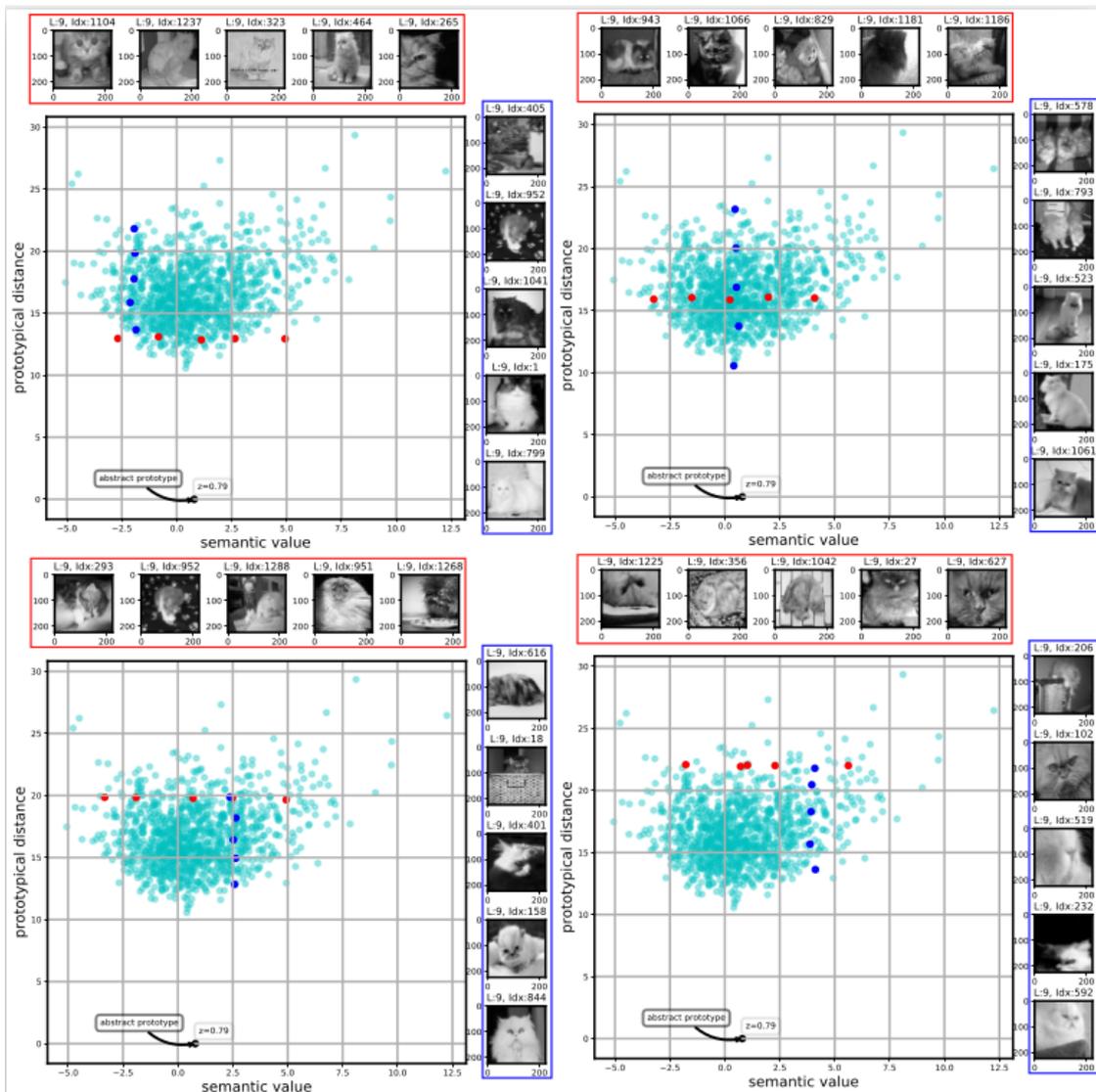


Fig. 11 *Typicality score analysis.* Objects images with same prototypical distance and different semantic values (in red) have similar visual representativeness within the category, and category members with different prototypical distance and same semantic value (in blue) are visually different. Also, we observe that object image visual representativeness (*typicality*) decreases as prototypical distance increases. Object image features were extracted with VGG16 model.

We observed what is the image visual information behavior when one of those semantic variables (*semantic value* and *prototypical distance*) change. We kept constant the value of one variable, and then, we analyzed the visual representativeness of the corresponding object images when the value of another variable increase. Figure 11 shows an example of this experiment within the Persian-cat ImageNet category. Note how for a fixed *prototypical distance* (*elements in red*), the semantic value variation does not generate significant changes in image visual representativeness (*typicality*) within the category. In contrast, for a fixed semantic value (*elements in blue*), the prototypical dis-

tance variation generates typicality ordered changes in the image’s visual information. We observed that when prototypical distance increases, object image visual typicality decreases. In contrast, the experiments did not allow to generalize a behavior pattern between semantic value and image typicality.

Based on the results of our experiments, we assumed that our semantic prototype representation correctly captures the central semantic meaning of images categories. Even with different CNN models and images datasets, our CPM model organizes the internal category structure following a prototypical organization of category members. Besides, we showed that

our *prototypical distance* influences elements arrangement around the category semantic prototype. Since our prototypical distance is a metric in CNN-feature domain, our semantic distance can be used as object image typicality score within the category (*typicality score* (o) = $1/\delta(o, P_i)$).

5.3 Global Semantic Descriptor based on Prototypes

5.3.1 Descriptor Configuration

By construction, the dimensionality of our GSDP descriptor signature depends of the object image CNN-features dimensionality (image features extracted with CNN classification model used in background), and dimensions of auxiliary matrix ($\chi_{r \times r}$) used as parameter in our $f(x)$ transformation (see Figure 7 and Algorithms 2, 3). With higher auxiliary matrix dimensionality, smaller is our GSDP signature size; and vice versa.

Consequently, since we needed the size variation of image CNN-features to evaluate our prototype-based description model, we used different CNN models as images features extractors. CNN models selection criteria were based on trying to evaluate our semantic description approach in different contexts: image CNN-features with different sizes, CNN models with varied architecture and depth, and image datasets of diverse nature (image resolution, image type, etc.). Also, for each CNN model used, we configured (using the auxiliary-matrix parameter) our semantic descriptor to return GSDP-signatures with noticeably different dimensionality.

Table 2 presents details of GSDP descriptor settings used to construct each semantic GSDP signature evaluated in our experiments. For each CNN-model used, we exhibit the CNN-feature status at each step of the workflow of our dimensionality reduction function ($f(x)$) (see Figure 7). Table 2 shows the CNN classification models used as feature extractor; image CNN-feature length ($|F|$); new CNN-feature shape ($F_{p \times q}$) after apply Step 1 of our $f(x)$ transformation; auxiliary matrix dimension ($\chi_{r \times r}$) used as parameter (we used two different configurations for each CNN model); number of matrices that make up our semantic gradient (g^{jk}); dimensionality of intermediary feature constructed with our $f(x)$ transformation ($|f(x)|$); and final length of GSDP signature ($|\psi|$) for each descriptor setting.

Note that our global semantic descriptor executes twice the $f(x)$ transformation to reduce the dimensionality of *semantic meaning* and *semantic difference* representations (see Algorithm 2). Consequently, GSDP

Table 2 Available GSDP descriptor signature dimensions for each CNN classification model used as features extractor.

CNN Model	$ F $	$F_{p \times q}$	$\chi_{r \times r}$	g^{jk}	$ f(x) $	$ \psi $
CNN-MNIST	128	16×8	8×8	2×1	16	32
			4×4	4×2	64	128
CNN-CIFAR	512	32×16	8×8	4×2	64	128
			4×4	8×4	256	512
VGG16	4096	64×64	16×16	4×4	128	256
			8×8	8×8	512	1024
ResNet50	2048	64×32	16×16	4×2	64	128
			8×8	8×4	256	512

signatures dimensionality is two times the dimensionality of $f(x)$ transformation features.

5.3.2 Signature Semantic Information

The experiments in the image CNN-features domain showed that *object semantic value* and *prototypical distance* organize all category members prototypically in a specific (and unique) position within the category semantic structure. The key idea behind our GSDP semantic descriptor is to encapsulate, in a vector representation, the same semantic interpretation –of image object features– captured by our CPM model. In this section, we show that our GSDP descriptor encodes and preserves the semantic information contained in an object features (semantic value and prototypical distance) used by our CPM model for semantic interpretation of object image. Also, we show how retrieving from GSDP descriptor signatures that semantic information and reconstructing the prototypical organization of object category achieved in the image CNN-features domain.

Let be (ψ_{c_i}, L_1) the metric space of object descriptor signatures. Descriptor properties 1 and 2 allow to easily recover the *object semantic value* and *prototypical distance* from GSDP descriptor signatures. Property 3 enables us to build descriptor signatures for abstract prototypes of categories. Similarly to ρ map, we can construct and show that map $\gamma : (\psi_{c_i}, L_1) \rightarrow (\mathbb{R}^2, L_1) \mid \gamma(\psi_o \in \psi_{c_i}) = p(\sum_0^{|\psi|/2} \psi_o, \sum_{|\psi|/2}^{|\psi|} \psi_o) = p(z_o, \delta(o, P_i))$ is continuous. Hence, we can map all category descriptor signatures to (\mathbb{R}^2, L_1) metric space using γ function.

With γ map approach, we can reproduce the same semantic analysis performed in CNN-feature space. Experiments showed that the category prototypical organization achieved in (\mathbb{R}^2, L_1) metric space is identical regardless of which γ map (to descriptor signature domain) or ρ map (to CNN-feature domain) function is used (e.g. Figure 10 and 11). Consequently, the behavior observed in (\mathbb{R}^2, L_1) metric space is equivalent to the behavior in feature metric space (F_{c_i}, δ) and descriptor signatures metric space (ψ_{c_i}, L_1) . This means that our

GSDP descriptor signature preserves, in its taxonomy, the same semantic information used by our CPM model to interpret object image CNN-features (*semantic value* and *prototypical distance*).

5.3.3 Signature Taxonomies

By definition, our GSDP descriptor uses category semantic prototypes as semantic distinctiveness generator of category members signatures. Elements with similar *semantic meanings* and that sharing similar *semantic differences* with the abstract prototype, will have similar GSDP semantic signatures. That is, since *abstract prototype* can be understood as a DNA chain that stands for the typical CNN-features of category members, the *abstract prototype signature* can be understood as a number distribution (or smaller DNA chain signature) that stands for category members signatures.

Figure 12 shows an example of the signatures taxonomies constructed with our GSDP semantic descriptor. We showed GSDP signatures constructed using CNN - MNIST model as features extractor of MNIST dataset images (signatures size = 32 since we used the GSDP minimal setting (see Table 2)). Also, we showed the structural polymorphism property (Property 3) of our GSDP descriptor to construct signatures for the *central semantic meaning* (abstract prototype) and category members. With our approach, category members will have semantic signatures with a similar representation of category abstract prototype signature. Notice that very typical category elements will have descriptor signatures similar to the abstract prototype signature, and elements that do not belong to the category will have a quite different GSDP signature.

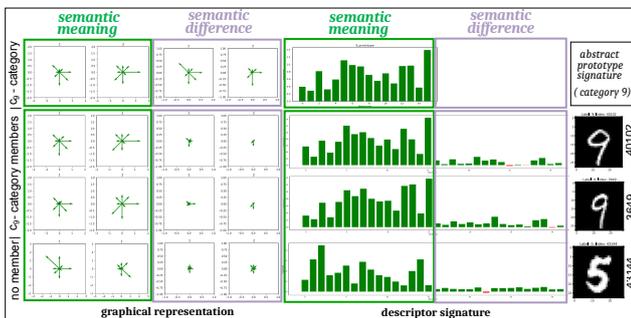


Fig. 12 *Semantic signature taxonomies.* Figure shows an example of semantic signatures constructed with our GSDP descriptor for c_9 -category in MNIST dataset. We show the abstract prototype signature, descriptor signatures examples of two c_9 -category members and a member that does not belong to c_9 -category.

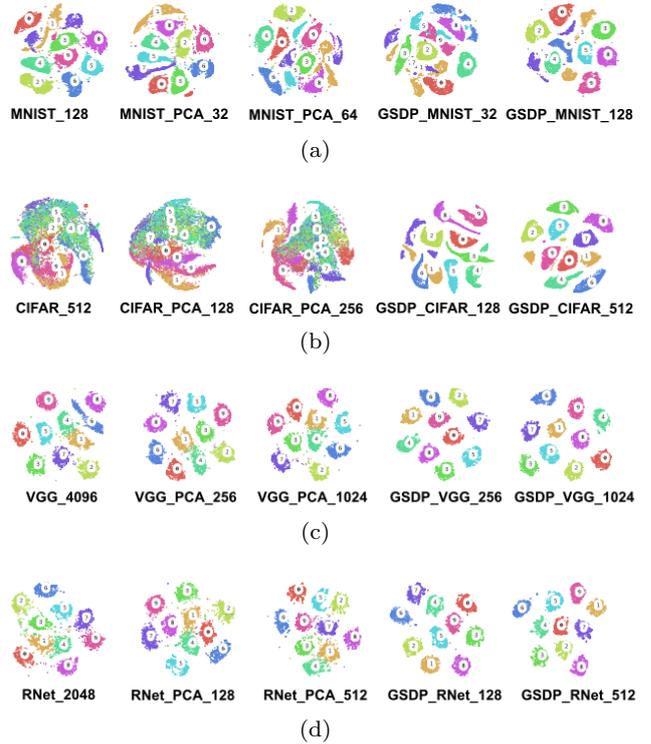


Fig. 13 *t-SNE visualization.* a) t-SNE visualization of features constructed with CNN-MNIST model in MNIST dataset; b) t-SNE visualizations of features constructed with CNN-CIFAR model in CIFAR10 dataset; c, d) t-SNE visualizations of first 10 categories of ImageNet dataset using features constructed with VGG16 and ResNet50 models, respectively. Each feature length was placed in the corresponding caption.

Our GSDP descriptor attempts to build – using our semantic prototype representation – a specific signature distribution for each object image category. Figure 10 and 11 show that category elements can be grouped, based on the meaning captured by our CPM model, by their family resemblance within the object category. However, this does not mean that in m -dimensional image features space, there are no elements of other categories in the neighborhood of a specific element. Since t-SNE algorithm (Maaten and Hinton 2008) can preserve the local structure, we used t-SNE to analyze the element neighborhood in m -dimensional space. Maaten and Hinton (2008) exposed that points which are close to one another in the high-dimensional dataset will tend to be close to one another in the t-SNE low-dimensional map.

We analyzed the discriminative power and t-SNE visualization performance of our GSDP semantic image representation *versus* features extracted using CNN classification models. For each CNN-model used as background by our GSDP descriptor, we compared the t-SNE performance of features family built with each

CNN model. We performed the t-SNE visualization experiment for features-family constituted by CNN features, corresponding GSDP semantic signatures, and reduced PCA versions of CNN-features (we reduced CNN-features to same GSDP feature dimensions).

Figure 13 shows the performance of t-SNE algorithm with each features-family in several image datasets using Euclidean distance as similarity measure and 50 as perplexity value. Note how GSDP representations achieved the best performance on each features-family. We observed that our GSDP object image representations are compactly grouped and have greater separation between categories than those t-SNE clustering built with high dimensionality features of CNN models (and its correspond PCA-reduced versions). Therefore, we can assume that our global semantic descriptor can construct object category representations distribution with the ability to maximize inter-class elements differences and minimize the intra-class differences. That is, with our approach, elements in each category must be as similar as possible, and elements in different groups must be as different as possible.

5.3.4 Performance Evaluation

Clustering

Yang et al. (2016) showed that when the image features representations achieve good metrics in image clustering task, it can generalize well when transferred to other tasks. Based on these assumptions, we evaluated our semantic GSDP encoding to verify its usefulness and suitability in image clustering task. We evaluated our GSDP descriptor (version based in VGG16 and ResNet50 classification models) performance in clustering task with ImageNet dataset image. We compared our GSDP representation performance against the following image global descriptors: GIST Oliva and Torralba (2001), LBP Ojala et al. (2002), HOG Dalal and Triggs (2005), Color64 Li (2007), Color_Hist Song et al. (2004), Hu_H_CH Haralick et al. (1973); Hu (1962); Song et al. (2004), VGG16 features and ResNet50 features (and its correspond PCA-reduced versions).

We used K-Means algorithm for clustering 50,000 images (500×category) of first 100 ImageNet dataset categories. The selection criteria of the K-Means algorithm is based on some similarities of the K-Means method with our image semantic representation approach. K-Means method minimizes the sum of squared errors between data points and their nearest cluster centers. This approach has similarities with our GSDP representation since GSDP signatures were constructed

Table 3 K-Means cluster metrics for each evaluated global image representation. Screenshot of K-Means measures for first 20 ImageNet categories (20 clusters). We show Homogeneity (H), Completeness (C), V-measure (V), Adjusted Rand Index (ARI) and Adjusted Mutual Information (AMI) clustering measures. We show in bold the best performance.

Descriptor	Size	FPS	Metrics Scores				
			H	C	V	ARI	AMI
GIST	960	0.82	0.05	0.05	0.05	0.01	0.05
LBP	512	0.72	0.02	0.03	0.03	0.01	0.02
HOG	1960	33	0.04	0.04	0.04	0.01	0.03
Color64	64	8	0.12	0.12	0.12	0.04	0.11
Color_Hist	512	26	0.08	0.08	0.08	0.03	0.07
Hu_H_CH	532	6.9	0.04	0.04	0.04	0.01	0.02
VGG16	4096	15	0.87	0.88	0.88	0.78	0.87
VGG_PCA_256	256	12.5	0.89	0.90	0.89	0.82	0.89
GSDP_VGG_256	256	12.8	0.97	0.99	0.98	0.93	0.97
VGG_PCA_1024	1024	12.5	0.89	0.89	0.89	0.81	0.89
GSDP_VGG_1024	1024	11.6	0.94	0.98	0.96	0.84	0.94
ResNet50	2048	10.6	0.88	0.90	0.89	0.78	0.88
RNet_PCA_128	128	12.5	0.88	0.88	0.88	0.81	0.88
GSDP_RNet_128	128	9.6	0.97	0.98	0.98	0.93	0.97
RNet_PCA_512	512	12.5	0.89	0.90	0.90	0.82	0.89
GSDP_RNet_512	512	9	0.91	0.97	0.94	0.73	0.91

to organize features categories using as category organization center the abstract prototype signature.

We evaluated each image representations performance in image clustering task comparing its K-Means clustering metrics (Homogeneity, Completeness, V-measure, Adjusted Rand Index, Adjusted Mutual Information). For each global image representation, the experiment was conducted incrementally, starting with 3 cluster (for 3 categories) and incrementing a category for each K-Means algorithm iteration. At the end of each K-Means execution, the clustering metrics were saved. The idea behind our clustering experiment was to evaluate each image representation performance as the amount and diversity of objects images increased.

Table 3 shows a screenshot of K-Means clustering metrics achieved by each global image descriptor for the first 20 ImageNet categories. Also, Table 3 shows features dimension and feature extraction velocity (frame per second - FPS) for each image representation approach. All experiments were performed on a standard computer, without the use of GPUs to be fair with handcraft features approaches.

Note that our object image semantic representation achieved the best performance among the image representations evaluated with the ImageNet images sample. Also, be mindful of our GSDP descriptor representation keeps the same semantic information used by our CPM model for VGG16 and ResNet50 features interpretation (See Section 5.3.2 and Figure 10), but with more discriminatory image representation and even lower feature dimension. Experiments showed that lowest dimensional GSDP representations obtained the best cost-benefit performance.

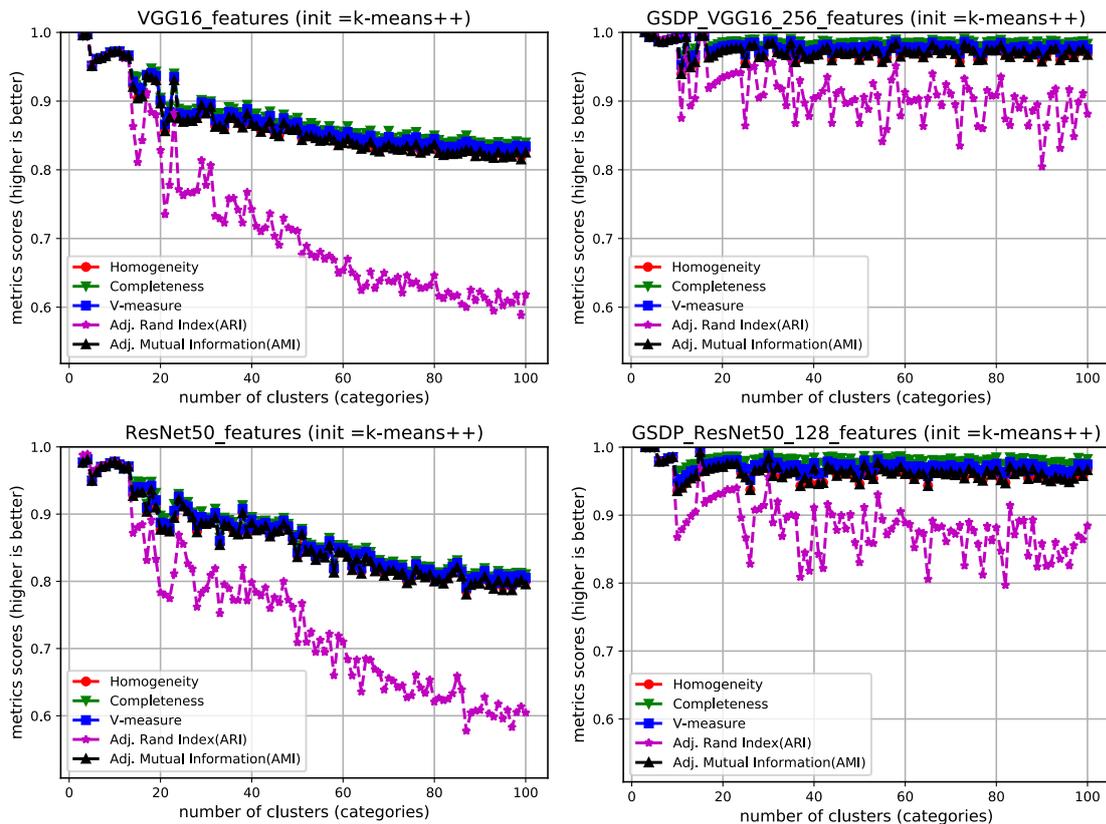


Fig. 14 History of K-Means metrics reached by each image feature representation in first 100 categories (98 K-Means iterations) of ImageNet dataset. We compared the performance of VGG16 and ResNet50 features (left) *versus* our GSDP descriptor signature (right) in clustering task.

Figure 14 shows K-Means metrics history for VGG16 and ResNet50 features representation against the correspond GSDP signatures. We showed K-Means metrics behavior for each image representation when the number of clusters increases (until 100 categories) in each K-Means algorithm execution. Experiments showed that as object images variety increased, K-Means clustering metrics related to CNN-features deteriorated significantly, while K-Means clustering metrics achieved by our image semantic encoding remain above 0.9. Results showed that our semantic descriptor encoding significantly outperforms others image global encodings in terms of cluster metrics.

Classification

To evaluate our image semantic encoding performance with supervised and unsupervised learning techniques, we also evaluated the performance of our GSDP representation in an image classification task.

Our GSDP descriptor, by construction, builds objects image representations based on object category predictions made by CNN model used as background

(See step depicted in Figure 6b and Algorithm 2 line 4). Consequently, a prediction error of CNN-classification models generates that our descriptor constructs an object’s image semantic representation using a wrong semantic prototype. This behavior is not problematic if we take into account that human beings will erroneously describe an object if it was previously wrong recognized.

In this experiment, we evaluated the performance of two GSDP semantic representations. Each GSDP semantic representation was constructed considering two different scenarios: *i*) images GSDP-signatures are made based on the prediction of CNN model used in the background (normal behavior of our GSDP descriptor); *ii*) images GSDP-signatures are made based on the prediction of an ideal classification model (100% accuracy) (a hypothetical behavior of our GSDP descriptor). We used as a prediction of an ideal classification model, the image category label annotated in ImageNet dataset. We conducted the experiment to analyze the possible performance of our GSDP representation if the prototype selection error is zero (prediction error of CNN-model used in background).

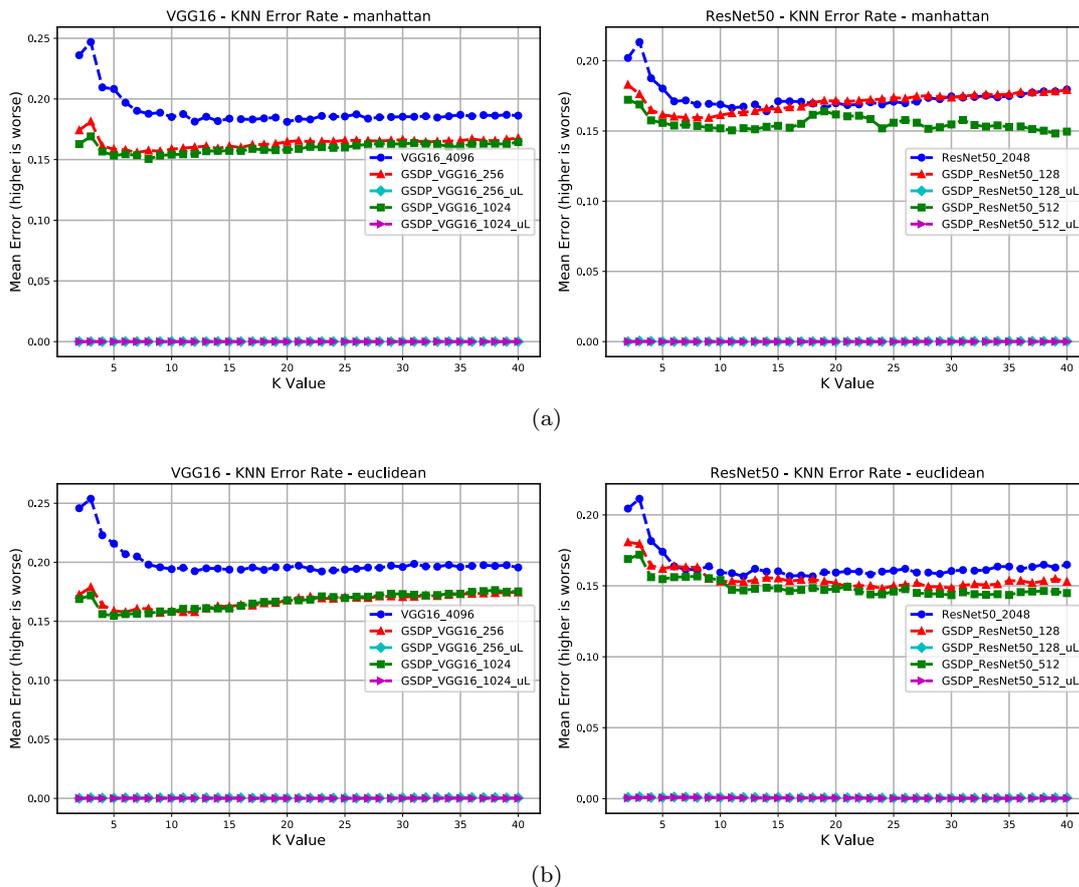


Fig. 15 KNN error rate reached by each image representation in the first 100 categories of ImageNet dataset. We varied the K-value of KNN algorithm to compare the performance of VGG16 and ResNet50 features *versus* our GSDP descriptor signature in image classification task using as feature similarity: a) Manhattan distance; b) Euclidean distance. Each feature-length was placed in the corresponding caption.

We performed our classification experiments using the KNN algorithm since, similar to t-SNE algorithm; elements are classified based on their local neighborhood. We analyzed the GSDP representation performance increasing the KNN algorithm parameter value (K neighbor) and using Euclidean and Manhattan distances as feature similarity measures.

Figure 15 shows KNN algorithm performance using VGG16 and ResNet50 representations against corresponding GSDP signatures constructed in those two scenarios. In the experiment, we used the same ImageNet images sample used for clustering task evaluation. Also, we varied the K-value to show that our GSDP encoding significantly outperforms VGG16 and ResNet50 encodings in the KNN classification task.

Experiments showed that our GSDP representation using the ResNet50 model reached a better performance than those constructed using the VGG16 model. Also, we observed that GSDP representations constructed using category labels (notated with `_uL` in Figure 15) are highly discriminative (mean error close to 0). Con-

sequently, we can conclude that our semantic encoding of objects substantially improves its performance – in classification task – as the accuracy of the CNN-classification model used in background increases.

Our experiments showed that our image global semantic representation based on category prototypes could outperform other image global representations in some computer vision tasks. Also, note that our GSDP encoding can describe objects images while encapsulates in its image signature the object semantic information (object semantic value and object typicality score). Experiments showed that lowest dimensional GSDP representations (for each CNN model) were the ones that achieved the best size-performance trade-off.

6 Limitations and Future works

The proposed *prototype-based description model* was constructed strictly to describe objects images, not to describe scenes images. Note that even when our seman-

tic descriptor was evaluated in images that representing scenes, our GSDP descriptor can outperform other image global representations. The results achieved encourage us to evaluate the generalization ability of our object semantic representation in other computer vision tasks as image retrieval and scene understanding.

As future work, the interpretative criteria of human beings is necessary to construct images dataset with typicality annotations and conclude if our model can interpret objects images similar to human beings.

7 Conclusion

Motivated by how human beings represent and relate the meanings attributed to objects, this research was based on the Prototype Theory to propose semantic representations of object categories and object images. Specifically, in this paper we introduced and evaluated two models based on Prototype Theory foundations: *i*) a Computational Prototype Model (CPM) and *ii*) a Prototype-based Description Model.

We proposed the CPM model to represent the internal semantic structure of object categories. Experiments showed that our CPM model was able to encapsulate relevant features of objects category in our semantic prototype representation. Also, we showed that our semantic distance metric could simulate semantic relationships in terms of visual typicality, between category members. Our experiments showed that a relationship could be established between our semantic distance metric and object image visual representativeness. Expressly, our prototypical distance can be understood as the object image typicality score. That is, our CPM model can capture the object’s visual typicality and the central and peripheral meaning of objects categories.

Based on the CPM model results, we proposed a prototype-based description model that uses the CPM model main components (*semantic prototype + semantic distance metric*) to construct a semantic representation of object’s image. Our prototype-based description model uses semantic prototypes of the CPM model to build a discriminatory signature that semantically describes object images highlighting its most distinctive features within the category.

Our novel Global Semantic Descriptor based on Prototypes (GSDP)¹ introduces a new approach to the semantic description of object images. GSDP descriptor does not need to be trained, and it is easily adaptable to be used with any CNN-classification model. As

shown in the experiments in the ImageNet dataset with VGG16 and ResNet50 models, our global semantic descriptor is discriminative, small dimensioned, and encodes the semantic information of category members. We further showed that our GSDP object representation preserves in its taxonomy the object’s semantic meaning and the object typicality score.

Our Prototype-based Description Model proposes a starting point to introduce the theoretical foundation related to *the representation of semantic meaning and the learning of visual concepts* of the Prototype Theory in the CNN semantic descriptors family.

Acknowledgements This research was supported by funding from the Brazilian agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG).

References

- Abdi H, Williams LJ (2010) Principal component analysis. Wiley interdisciplinary reviews: computational statistics 2(4):433–459
- Atkinson RC, Shiffrin RM (1968) Human memory: A proposed system and its control processes. Psychology of learning and motivation 2:89–195
- Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (surf). Computer Vision and Image Understanding (CVIU) 110(3):346–359
- Binder JR, Conant LL, Humphries CJ, Fernandino L, Simons SB, Aguilar M, Desai RH (2016) Toward a brain-based componential semantic representation. Cognitive Neuropsychology 33(3-4):130–174
- Bristow H, Valmadre J, Lucey S (2015) Dense semantic correspondence where every pixel is a classifier. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp 4024–4031
- Cadiou CF, Hong H, Yamins DL, Pinto N, Ardila D, Solomon EA, Majaj NJ, DiCarlo JJ (2014) Deep neural networks rival the representation of primate it cortex for core visual object recognition. PLoS Computational Biology 10(12):e1003963
- Chebyshev P (1867) Des valeurs moyennes. Journal de Mathématiques pures et Appliquées 12:177–184
- Cichy RM, Khosla A, Pantazis D, Oliva A (2017) Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. NeuroImage 153:346–358
- Collins JA, Curby KM (2013) Conceptual knowledge attenuates viewpoint dependency in visual object recognition. Visual Cognition 21(8):945–960
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, vol 1, pp 886–893
- Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T (2014) Decaf: A deep convolutional activation feature for generic visual recognition. In: Proceed-

¹ All source code, prototypes datasets, and GSDP tool tutorials are publicly available in our lab’s Github: <https://github.com/verlab/gsdp>.

- ings of the International Conference on Machine Learning (ICML), pp 647–655
- Estes W (1986) Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General* 115(2):155
- Fromkin V, Rodman R, Hyams N (2018) *An introduction to language*. Cengage Learning
- Fuster JM (1997) Network memory. *Trends in Neurosciences* 20(10):451–459
- Geeraerts D (2010) *Theories of lexical semantics*. Oxford University Press
- Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS (2016) Deep learning for visual understanding: A review. *Neurocomputing* 187:27–48
- Han K, Rezende RS, Ham B, Wong KYK, Cho M, Schmid C, Ponce J (2017) Snet: Learning semantic correspondence. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 1831–1840
- Han X, Leung T, Jia Y, Sukthankar R, Berg AC (2015) Matchnet: Unifying feature and metric learning for patch-based matching. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 3279–3286
- Haralick RM, Shanmugam K, et al. (1973) Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* 6(6):610–621
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 770–778
- Homa D, Vosburgh R (1976) Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory* 2(3):322
- Hu MK (1962) Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* 8(2):179–187
- Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Computational Biology* 10(11):e1003915
- Kim S, Min D, Ham B, Lin S, Sohn K (2018) Fcss: Fully convolutional self-similarity for dense semantic correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*
- Krizhevsky A, Hinton G (2010) Convolutional deep belief networks on cifar-10. Unpublished manuscript 40
- Lake B, Zaremba W, Fergus R, Gureckis T (2015) Deep neural networks predict category typicality ratings for images. In: *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, Cognitive Science Society
- Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324, DOI 10.1109/5.726791
- Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorization. In: *Advances in Neural Information Processing Systems (NIPS)*, pp 556–562
- Li M (2007) Texture moment for content-based image retrieval. In: *2007 IEEE International Conference on Multimedia and Expo, IEEE*, pp 508–511
- Li O, Liu H, Chen C, Rudin C (2018) Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions. In: *Thirty-Second AAAI Conference on Artificial Intelligence*
- Lin K, Lu J, Chen CS, Zhou J (2016) Learning compact binary descriptors with unsupervised deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 1183–1192
- Liu C, Yuen J, Torralba A (2011) Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 33(5):978–994
- Long JL, Zhang N, Darrell T (2014) Do convnets learn correspondence? In: *Advances in Neural Information Processing Systems (NIPS)*, pp 1601–1609
- Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)* 60(2):91–110
- Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. *Journal of Machine Learning Research* 9(Nov):2579–2605
- Martin A (2007) The representation of object concepts in the brain. *Annual Review of Psychology* 58:25–45
- McRae K, Jones M (2013) Semantic memory. *The Oxford handbook of Cognitive Psychology* 206
- Medin DL, Schaffer MM (1978) Context theory of classification learning. *Psychological review* 85(3):207
- Minda JP, Smith JD (2001) Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(3):775
- Minda JP, Smith JD (2002) Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(2):275
- Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115(1):39
- Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 24(7):971–987
- Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision (IJCV)* 42(3):145–175
- Reed SK (1972) Pattern recognition and categorization. *Cognitive Psychology* 3(3):382–407
- Rocco I, Arandjelović R, Sivic J (2018) End-to-end weakly-supervised semantic alignment. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
- Rosch E (1975) Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104(3):192
- Rosch E (1978) Principles of categorization. In: Rosch E, Lloyd BB (eds) *Cognition and Categorization*, Hillsdale, NJ:Lawrence Erlbaum Associates, pp 27–48
- Rosch E (1988) Coherences and categorization: A historical view. *The development of language and language researchers: Essays in honor of Roger Brown* pp 373–392
- Rosch E, Lloyd BB (1978) *Cognition and categorization*, vol 1. Lawrence Erlbaum Associates Hillsdale, NJ
- Rosch E, Mervis CB (1975) Family resemblances: Studies in the internal structure of categories. *Cognitive psychology* 7(4):573–605
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115(3):211–252, DOI 10.1007/s11263-015-0816-y

- Saw JG, Yang MC, Mo TC (1984) Chebyshev inequality with estimated mean and variance. *The American Statistician* 38(2):130–132
- Simo-Serra E, Trulls E, Ferraz L, Kokkinos I, Fua P, Moreno-Noguer F (2015) Discriminative learning of deep convolutional feature point descriptors. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp 118–126
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556*
- Simonyan K, Vedaldi A, Zisserman A (2014) Learning local feature descriptors using convex optimisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 36(8):1573–1585
- Song Yj, Park Wb, Kim Dw, Ahn Jh (2004) Content-based image retrieval using new color histogram. In: *Intelligent Signal Processing and Communication Systems, 2004. IS-PACS 2004. Proceedings of 2004 International Symposium on, IEEE*, pp 609–611
- Stellato B, Van Parys BP, Goulart PJ (2017) Multivariate chebyshev inequality with estimated mean and variance. *The American Statistician* 71(2):123–127
- Sternberg RJ, Sternberg K (2016) *Cognitive psychology*. Nelson Education
- Strecha C, Bronstein A, Bronstein M, Fua P (2012) Ldash: Improved matching with smaller descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 34(1):66–78
- Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-First AAAI Conference on Artificial Intelligence*, pp 4278–4284
- Thompson-Schill SL (2003) Neuroimaging studies of semantic memory: inferring how from where. *Neuropsychologia* 41(3):280–292
- Tola E, Lepetit V, Fua P (2008) A fast local descriptor for dense matching. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp 1–8
- Tulving E (2007) Coding and representation: searching for a home in the brain. *Science of Memory: Concepts* pp 65–68
- Wohllhart P, Köstinger M, Donoser M, Roth PM, Bischof H (2013) Optimizing 1-nearest prototype classifiers. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp 460–467
- Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences* 111(23):8619–8624
- Yang H, Lin WY, Lu J (2014) Daisy filter flow: A generalized discrete approach to dense correspondences. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 3406–3413
- Yang J, Parikh D, Batra D (2016) Joint unsupervised learning of deep representations and image clusters. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 5147–5156
- Yi KM, Trulls E, Lepetit V, Fua P (2016) Lift: Learned invariant feature transform. In: *Proceedings of the of the European Conference on Computer Vision (ECCV)*, Springer, pp 467–483
- Zagoruyko S, Komodakis N (2015) Learning to compare image patches via convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 4353–4361
- Zaki SR, Nosofsky RM, Stanton RD, Cohen AL (2003) Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory and Cognition* 29(6):1160–1173