Semantic Description of Objects in Images Based on Prototype Theory

Omar Vidal Pino, Erickson R. Nascimento, and Mario F. M. Campos
Department of Computer Science
Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
Email: {ovidalp, erickson, mario}@dcc.ufmg.br

***Abstract*—This research aims to build a model for the semantic description of objects based on visual features extracted from images. We introduce a novel semantic description approach inspired by the Prototype Theory. Inspired by the human approach used to represent categories, we propose a novel Computational Prototype Model (CPM) that encodes and stores the object's image category's central semantic meaning: the semantic prototype. Our CPM model represents and constructs the semantic prototypes of object categories using Convolutional Neural Networks (CNN). The proposed Prototype-based Description Model uses the CPM model to describe an object highlighting its most distinctive features within the category. Our Global Semantic Descriptor (GSDP) builds discriminative, low-dimensional, and semantically interpretable signatures that encode the objects' semantic information using the constructed semantic prototypes. It uses the proposed Prototypical Similarity Layer (PS-Layer) to retrieve the category prototype using the principle of categorization based on prototypes. Using different datasets, we show in our experiments that: *i)* the proposed CPM model successfully simulates the internal semantic structure of the categories; *ii)* the proposed semantic distance metric can be understood as the object typicality score within a category; *iii)* our semantic classification method based on prototypes can improve the performance and interpretation of CNN classification models; *iv)* our semantic descriptor encoding significantly outperforms others state-of-the-art image global encoding in clustering and classification tasks.**

I. Introduction

For several years, the fields of Computer Vision and Machine Learning have tried to build pattern recognition methods with a similar performance of a human being for visual information understanding. 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 [1]. The advent of Convolutional Neural Networks (CNN) outperformed the traditional methods used for image feature representation [2]–[4] and it enabled to achieve a visual recognition model with similar behavior of *human semantic memory* [5] for classification tasks [6]–[8]. CNN's success sparked the tendency of images semantic processing with deep-learning techniques. Representations of image features extracted using deep classification models [6]–[8], or using CNN-descriptors are commonly referred as *semantic feature* or *semantic signature*.

Semantic feature and *Semantic Meaning* terms have been extensively studied in the field of linguistic semantic. While a semantic feature is defined as the representation of the basic conceptual components of the meaning of any lexical item [9], according to Rosch [10], the representation of *semantic meaning* is related to the *category prototype*, particularly to those categories naming natural objects. In her seminal work [10], Rosch introduced the concept of *semantic prototype* and presented a deep analysis of the semantic structure in the meaning of words. Although state-of-the-art methods have achieved surprising results, there are still many challenges to simulate the discriminative and abstraction power of human semantic memory to represent the semantics.

In this research[1], we rely on cognitive semantic studies related to the Prototype Theory [10]–[12] for modeling the *central semantic meaning* of objects categories: the prototype. The observations on the Prototype Theory raise the following questions: i) How to describe and stand for objects images, semantically? ii) 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? iii) How can the category prototype be included in the object global semantic description and classification tasks?

We address these questions motivated by the human's approach to describe objects globally. Humans use 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). Fig. 1 illustrates the intuition and principal concepts of our prototype-based description model. Our approach's main idea is to use the quality of features extracted with CNN-classification models both to represent the central semantic meaning of a specific category and learn the object distinctiveness within the category.

More specifically, our main contributions in this work 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 (prototypes); 2) a *semantic distance metric* in object image CNN features domain, which can be understood as a measure of object

[1]This work relates to a Ph.D. thesis. See the project page: https://www.verlab.dcc.ufmg.br/global-semantic-description.

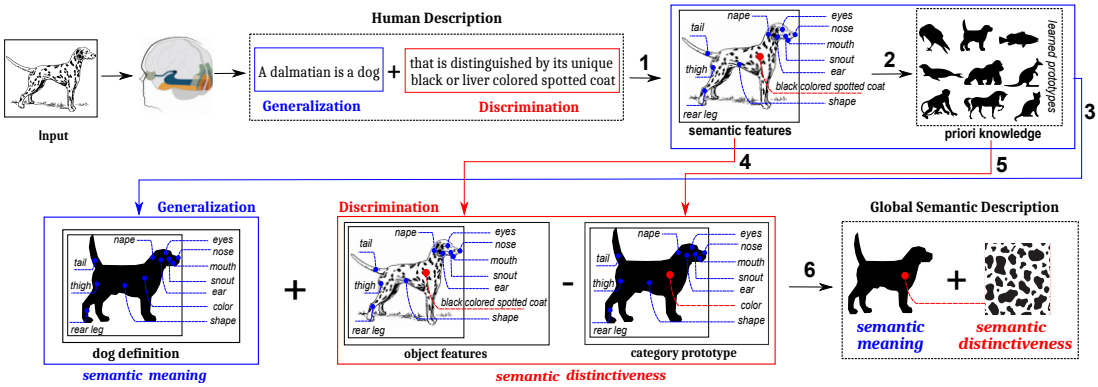

Fig. 1. **Motivation and Concepts**. Schematic of our prototype-based description model. The human visual system can observe an object and build an object semantic description that highlighting their most distinctive features within the object category. We propose a prototype-based model to simulate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

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; 4) a *prototype-based semantic classification model* for semantic classification of objects images based on prototypes. We propose a *Prototypical Similarity Layer (PS-Layer)* that classifies objects according to its similarity concerning our prototypes encoding.

II. Related Works

A. CNN descriptors

CNNs provide outstanding performance in image semantic processing tasks, and descriptors extracted using CNN techniques have outperformed the best techniques based on carefully hand-crafted image features [2]–[4]. CNN descriptor models differ among themselves on how to compute the image representation in their deep architectures [13]–[15], similarity functions learning [14], [16], and its features extraction methods [7], [8], [14]. Initially, CNN descriptor models were more oriented toward achieving discriminatory features than representing the image's semantic information. Still, some works [17], [18] use the robustness of CNN-models for training semantic descriptors architectures to address the problem of semantic correspondence [19]. In general, CNN descriptors and semantic descriptors are trained to learn semantic representations with different approaches and architectures. CNN descriptors and semantic descriptors effectively learn their image representations, but it is still unknown how they encapsulate semantics. Nevertheless, *none* of these CNN-feature description approaches codify the representation of the visual information based on the theoretical foundation of Cognitive Science to represent the *semantic meaning*. In contrast, in this thesis, we introduce a novel image semantic description approach based on the foundation of Prototype Theory to represent the meaning of an object's image.

B. Prototype Theory

The Prototype Theory [10]–[12], [20]–[22] analyzes the internal semantic 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 [10], [11] obtained evidence that human beings store first the *semantic meaning of category* based on degrees of representativeness (*typicity*) of category members, and then its specificities. The *category prototype* was formally defined as the clear central members (typical members) of a category [10], [12], [21]. Rosch [10], [11], [20] showed that human beings store the category knowledge as a semantic organization around the category prototype (*prototypicality organization* phenomenon [11], [21], [23]). Finally, object categorization is obtained based on the similarity of a new exemplar with the learned categories prototypes [11], [20]. For Geeraerts [21], four characteristics are frequently mentioned as typical of prototypicality in prototypical categories [10], [11], [21]: *i) extensional non-equality*; *ii) extensional non-discreteness*; *iii) intensional non-equality*; and *iv) intensional non-discreteness*. The *prototypicality effects* surmise the importance of the distinction between *central* and *peripheral* meaning of the object categories [21]. In this thesis, we try to model the main concepts of the prototypicality effects.

C. The prototype in classification tasks

Learning Vector Quantization (LVQ) is a field started by the seminal work of Kohonen [24], in which the methods try to find optimal prototypes from the labeled data. This approach [24]–[26] divides the feature input space assigning data samples to a set of prototypes. LVQ models have the advantage that reduce the complexity of the classifier, but its performance strongly depend on the correct prototypes choice [26]. The learning method always tries to move the prototypes near to training samples of the same category and away from other categories. Recent works use prototypes to improve the classification on zero-shot learning [27] and few-

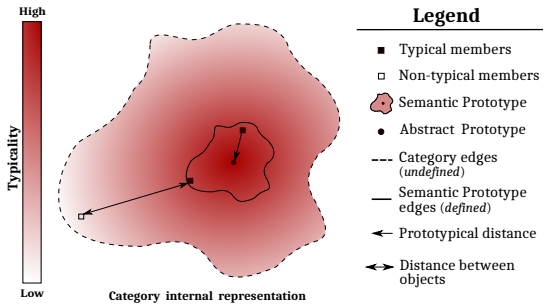

Fig. 2. **Category internal structure**. Figure shows our expected semantic representation of category internal structure. Also it shows the principal definitions and constraints of our Computational Prototype Model.

shot learning [28]. In general, works that use prototypes as the core of its classification learning process: use a set of images templates as priori prototypes; learn or calculate the prototypes in a learned embedded space, and classify a new instance image based on the similarity between each instance-prototype pair computed with a distance metric (commonly euclidean distance or some learned metric). Although the prototype be considered (conceptually) the most typical member of the category, it is common that literature approaches do not introduce the *image typicality* in category learning processes.

III. Methodology

A. Computational Prototype Model

Based on our hypothesis for simulating the human behavior in object features description (see Fig. Figure 1), our proposal requires as *priori knowledge* the representation of objects categories prototypes.We proposed a mathematical framework, based on Prototype Theory foundations, to stand for the central semantic meaning of object image categories. Our *Computational Prototype Model* (CPM) is a set of definitions and constraints that allows us to interpret possible semantic associations between members within the internal category structure (See Fig. 2).

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

Definition 1. *Semantic prototype of c_i-category is the 3-tuple $P_i = (M_i, \Sigma_i, \Omega_i)$ where:* ***i)*** *$M_i = [\mu_{i1}, \mu_{i2}, ..., \mu_{im}]$ is a 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)*** *$\Sigma_i = [\sigma_{i1}, \sigma_{i2}, ..., \sigma_{im}]$ is a 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; and* ***iii)*** *$\Omega_i = [\omega_{i1}, \omega_{i2}, ..., \omega_{im}]$ is a nonempty m-dimensional vector, where ω_{ij} is the relevance value of j-th feature for the category $c_i \in C$.*

Algorithm 1 Semantic Prototype Construction
Input: CNN-Model Λ, Object Dataset O, Category c_i
Output: Category Prototype (P_i)
Initialization :
$O_{c_i} \leftarrow \{o \in O : category(o) = c_i\}$
$features\_block \leftarrow \{\}$
$threshold \leftarrow 0.99$
for $o \in O_{c_i}$ **do**
  F_o, $typicality\_score \leftarrow \Lambda.features\_of(o)$
  if $(typicality\_score \geq threshold)$ **then**
    $features\_block \leftarrow features\_block \cup F_o$
  end if
end for
$\Omega_i, b_i \leftarrow \Lambda.softmax\_weight\_learned\_of(c_i)$
$M_i, \Sigma_i \leftarrow compute\_stats(features\_block)$
return $(M_i, \Sigma_i, \Omega_i, b_i)$

Definition 2. *Distance between objects.*[2] *Let $o_1, o_2 \in O_{c_i}$ be objects of i-th category $c_i \in C$; $\mathrm{F}_{o_1}, \mathrm{F}_{o_2}$ the features of objects o_1, o_2 respectively. We defined the semantic distance between objects o_1 and o_2 as: $\delta(o_1, o_2) = \sum_{j=1}^{m} |\omega_{ij}| \left| f_j^1 - f_j^2 \right|$, where $\omega_{ij} \in \Omega_i$, $f_j^1 \in F_{o_1}$ and $f_j^2 \in F_{o_2}$.*

Definition 3. *Prototypical distance.*[3] *Let $\mathrm{o} \in O_{c_i}$ be an object of i-th category $c_i \in C$, F_o the features of the 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: $\delta(\mathrm{o}, P_i) = \sum_{j=1}^{m} |\omega_{ij}| \left| f_j - \mu_{ij} \right|$, where $\omega_{ij} \in \Omega_i$, $\mu_{ij} \in M_i$, $f_j \in F_o$; and $M_i, \Omega_i \in P_i$.*

Definition 4. *Semantic prototype edges.*[4] *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 from only typical objects of c_i-category, and $F_o \subseteq E$ the features of a typical object $o \in O_{c_i}$. We weakly defined as edges of our semantic prototype P_i, the threshold vector $\vec{\lambda_i} = [\lambda_{i1}, \lambda_{i2}, ..., \lambda_{im}]$ that meets the expression: $\Pr(|f_j - \mu_{ij}| \geq \lambda_{ij}\sigma_{ij}) \leq \min\left(1, \frac{1}{\lambda_{ij}^2}\right)$, where $f_j \in F_o$, $\mu_{ij} \in M_i$ and $\sigma_{ij} \in \Sigma_i$.*

B. 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 Fig. 1. We lay hold of the theoretical foundations related to the representation of semantic meaning [5], [32], [33] to model an object semantic meaning representation. We define the *object semantic value* $z = \sum_m \omega_{ij} f_j + b_i$ to be the same value used to object categorization in softmax layer of CNN-classification models [34]. Hence, we assume 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*.

[2]Based on *the psychological distances between two stimuli* proposed in Generalized Context Model (GCM) [22], [29].
[3]Based on the distance of *Multiplicative Prototype Model* (MPM) [30].
[4]Based on Multivariate Chebyshev inequality constraints [31].

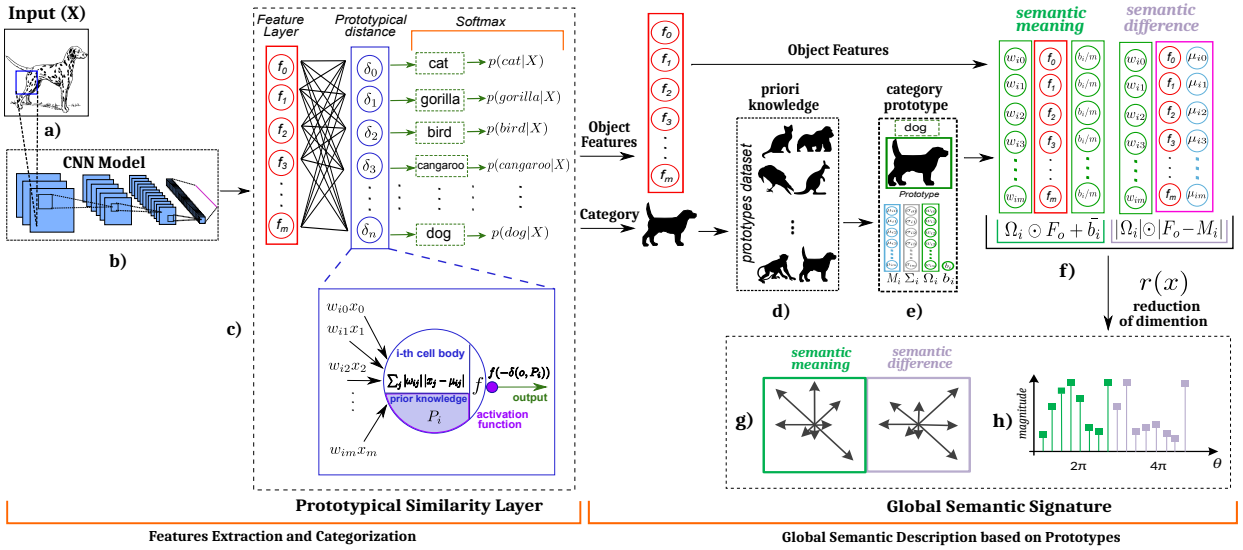

Fig. 3. **Methodology Overview**. Set of steps to transform the visual information received as input into a Global Semantic Descriptor signature. The methodology workflow can be divided into two main stages: ***1)*** *Feature extraction and Categorization* and ***2)*** *transformation of CNN-object features into our Global Semantic Signature*. *a*) input image; *b)-c)* features extraction and classification using a pre-trained CNN-classification model. Our Prototypical Similarity Layer (PS-Layer) is used to convert a common CNN-model into a prototype-based CNN-classification model; *d)* prototype dataset; *e)* category prototype selection; *f)* global semantic description of object using category prototype; *g)* graphic representation of our descriptor signature resulting from the dimensionality reduction function ($r(x)$); and *h)* Global Semantic Signature.

We stand for the semantic distinctiveness of an object for a specific c_i-category as the semantic discrepancy between object features and features of the c_i-category prototype. Consequently, we assume 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 (See Definition 3).

Fig. 3 shows an overview of our novel prototype-based description model. Our *Global Semantic Descriptor based on Prototypes (GSDP)* requires the *priori knowledge* of each category prototype (pre-computed off-line using Algorithm 1). After feature extraction and categorization processes (Fig. 3a-c), we use the corresponding category prototype for semantic description of object features. We show in Fig. 3f) the steps to introduce the category prototype into the global semantic description of object's features. A drawback of our object semantic representation (Fig. 3f) is having high dimensionality, since it is based on *semantic meaning vector* ($\vec{z}$) and *semantic difference vector* ($\vec{\delta}$). We proposed the transformation function $r(x)$ [34] to compress our global semantic representation of the object's features (Fig. 3f) in a low dimensional global semantic signature (Fig. 3g). Fig. 3 details the main steps of our description approach; note that we follow the same work-flow of human description hypothesis depicted in Fig. 1.

C. Prototype-based Semantic Classification

We present how to introduce our CPM framework to simulate the prototype-based concept of categorization of Prototype Theory (see work-flow depicted in Figure 1 steps 1-3)). Figure 3c) presents the internal structure of our *Prototypical Similarity Layer (PS-Layer)*: *i)* we show how to use our PS-Layer in a common CNN-model; *ii)* we highlighted in purple the mathematical model of our PS-Layer neuron. Noting how the cell neuron body keeps, as priori knowledge, our $i-th$ category semantic prototype. The PS-Layer has many neurons as prototypes and categories (see Fig. 3c); and it uses as neuron output activation the object's semantic distinctiveness (our prototypical distance). Analogous to MPM model [22], [30], our PS-Layer computes the probability with which object $o \in O$ is classified into i-th category using the equation: $P(c_i|o) = S(o,P_i)^\gamma / \sum_{k=1}^{n} S(o,P_k)^\gamma$ where γ is the *response-scaling* parameter, and $S(o,P_i) = exp(-\alpha\delta(o,P_i))$ is the similarity between object $o \in O$ and the i-th prototype (P_i). Without loss of generality, and using the same MPM model assumptions [22], [30], we set at 1 the α and γ parameters. Consequently, classification probability of our PS-Layer can be rewritten as:

$$P(c_i|o) = \frac{\exp(-\delta(o,P_i))}{\sum_{k=1}^{n}\exp(-\delta(o,P_k))}, \tag{1}$$

where $\delta(o,P_i)$ is our prototypical distance. Note that our PS-Layer uses a softmax function over our *prototypical distance* as probability distribution: $P = softmax(-\vec{\delta}(o,P_k))$. To simplify our PS-Layer neuron gradient computation, we add several constraints: *i)* neuron weights must be non-negative ($\omega_{ij} \geq 0$); ii) L2-regularization is used to guarantee small weights values [22]. Consequently, since $\mu_{ij} \in M_i$ is a constant, our PS-Layer neuron gradient:

$$\frac{\partial\delta}{\partial\omega} = \begin{cases} \mu_i - x, & \text{if } \omega x - \omega\mu_i \geq 0 \\ x - \mu_i, & \text{if } \omega x - \omega\mu_i < 0. \end{cases} \tag{2}$$

is as simple as common CNN neuron gradient $\partial z/\partial\omega = x$. Then, the model that uses our PS-Layer can be trained using the same training conditions of a baseline CNN-model.

IV. Experiments and Results

A. Experimental Setup

1) ***Datasets****:* The off-line prototype computation process was conducted using MNIST [35], CIFAR10, CIFAR100 [36] and ImageNet [37] datasets. We evaluated our GSDP descriptor performance in ImageNet [37] and Coco [38] as real images datasets. For each image dataset, we used a CNN-classification model for feature extraction and classification (see Fig. 3a-c). Our PS-Layer performance was evaluated in MNIST [35], CIFAR10 and CIFAR100 [36] datasets.

2) ***Models****:* We evaluated our GSDP descriptor using CNN-models architectures based on *LeNet* [35] and *Deep Belief Network* [36] architectures for MNIST and CIFAR datasets, respectively. Also, we conducted experiments in ImageNet [37] and Coco [38] using VGG16 [7] and ResNet50 [6] models as background of our global semantic description model. Also, PS-Layer performance evaluation was conducted using some models architectures: *sMNIST* [35], *sCF10* [36], *sCF100* [36], *vggCF10* [39], *vggC100* [39].

B. Computational Prototype Model

There is no defined metric to quantify whether our CPM framework correctly captures the category semantic meaning. 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.

1) ***Semantic prototype encoding****:* We analyze the semantics behind of our semantic prototype representation. We conducted the hierarchical clustering of our categories semantic prototypes to illustrate the hierarchical semantic organization of a specific image dataset. Fig. 4 shows an example of a tree diagram (dendrogram) achieved by semantic prototypes computed in CIFAR10. Notice how our semantic categories representations partition the CIFAR10-dataset achieving a hierarchical semantic organization. For example, two macro-categories are clearly visible in Fig. 3: *animals* and *transport vehicles*. Note also how this last macro-category is also semantically interpreted by our representation as *non-ground vehicles* and *ground vehicles*.

2) ***Central-Peripheral meaning and Prototypical Organization****:* We conducted experiments to know what is the visual representativeness of category members closest and furthest

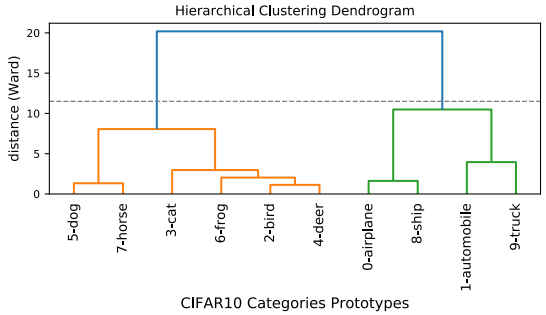

Fig. 4. **Hierarchical clustering of CIFAR10 semantic prototypes**.

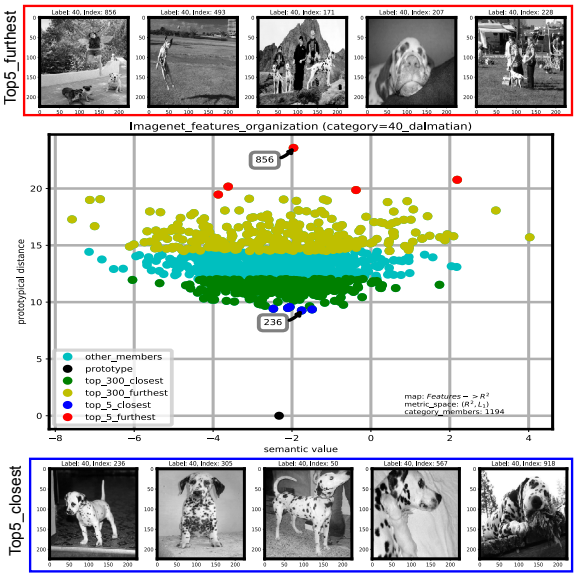

Fig. 5. **Prototypical organization within dalmatian category of ImageNet.** We represented category members using image features extracted with VGG16 model.

from the category semantic center (our abstract prototype). We proposed a function $\rho : F_{c_i} \rightarrow \mathbb{R}^2 \mid \rho(F_o) = p(z_o, \delta(o, P_i))$ [34] that maps image object features to $(\mathbb{R}^2, L1)$ metric space using its *semantic value* and its *prototypical distance*. Fig. 5 shows an example of the internal semantic structure captured by our CPM-model. Note that our CPM-model can recognize as most visually representative category-members (Top-5 closest members to semantic prototype (blue)), those objects' images that are easy recognized as it exhibits the category typical features. Also, we observed that elements identified as less representatives (Top-5 furthest elements (red)) are not easily recognized by human beings, although retaining some category features. Our approach allow to observe how visually relevant are those elements allocated by our CPM model in *center and periphery* of category.

3) ***Image Typicality Score****:* Lake *et al*. [40] showed that *semantic value* can be used as a signal for how typical an input image looks like. We conducted qualitative experiments to analyze how semantic value variations vs. prototypical distance can influence object image visual representativeness (typicality). In contrast to Lake *et al*. results, our experiments showed that using the *semantic value* as object typicality score can be problematic since objects with same semantic value do not imply same image typicality (*e.g.*, Fig. 5). We observed that when prototypical distance increases, object image visual typicality decreases (*typicality score* $(o) = 1/\delta(o, P_i)$). However, experiments did not allow us to generalize a behavior pattern between semantic value and image typicality.

C. Global Semantic Descriptor based on Prototypes

We evaluated our image semantic encoding performance with supervised (image classification) and unsupervised learning (clustering) techniques.We evaluated our GSDP descriptor performance in clustering task (comparing its K-Means

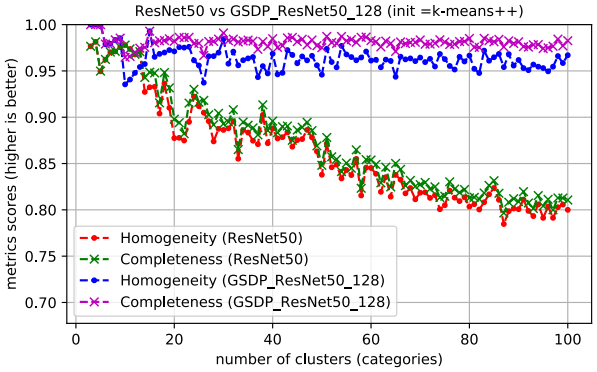

Fig. 6. **K-Means metrics on ImageNet**. History of K-Means metrics reached by ResNet50 features *versus* our GSDP representation in ImageNet dataset.

TABLE I
K-MEANS CLUSTER METRICS ACHIEVED FOR RESNET50 FEATURES *versus* OUR GSDP REPRESENTATION IN COCO DATASET.

Descriptor	Size	FPS	Metrics Scores				
			H	C	V	ARI	AMI
Deep Features Performance on Coco [38](CrossDataset)							
ResNet50 [6]	2048	10.6	0.29	0.36	0.32	0.17	0.31
ResNet50_PCA_128	128	12.5	0.32	0.34	0.33	0.17	0.31
ResNet50_PCA_512	512	12.5	0.34	0.35	0.34	0.20	0.33
GSDP_RNet_128 (our)	128	9.6	**0.43**	**0.69**	**0.53**	**0.16**	**0.52**
GSDP_RNet_512 (our)	512	9	0.34	0.47	0.40	0.09	0.39

clustering metrics) using ImageNet and Coco datasets. We compared our GSDP representation performance on ImageNet dataset against: ***1)*** traditional handcraft image global descriptors: GIST [41], LBP [42], HOG [43], Color64 [44], Color_Hist [45], Hu_H_CH [45]–[47]; ***2)*** deep learning images features trained on ImageNet: VGG16 features and ResNet50 features (and PCA-reduced versions).

Fig. 6 shows an example of K-Means metrics history achieved for ResNet50 features against our GSDP signatures. Experiments showed that as the data diversity of object's images increases, our semantic GSDP encoding significantly outperforms other image global encoding in terms of cluster metrics in ImageNet dataset. Also, we conducted the same experiments on Coco (cross-dataset) to evaluate the performance and generalization ability of each image representation on unseen data. Table I shows a screenshot with an example of K-Means clustering metrics achieved by each global image descriptor on the 18-th iteration of the experiments.

D. Prototype-based Classification Model

We evaluated the PS-Layer performance using as *baseline* models the CNN architecture: *sMNIST*, *sCF10*, *sCF100*, *vggCF10*, and *vggC100*. For each baseline model, we replaced the *softmax layer* with our PS-Layer. We trained the resulting *PS-Layer models* changing the weights initialization method: *fromscratch*, *freezing* and *pretrain*. For each initialization method we used two distance function inside the PS-Layer: *a) prototypical distance*; and *b) penalized prototypical distance* (we penalized peripheral elements using Def. 4 constraints). Consequently, we evaluated six PS-Layer model versions. Fig. 7 summarizes the performance of each PS-Layer model versions evaluated. Experiments showed that our PS-

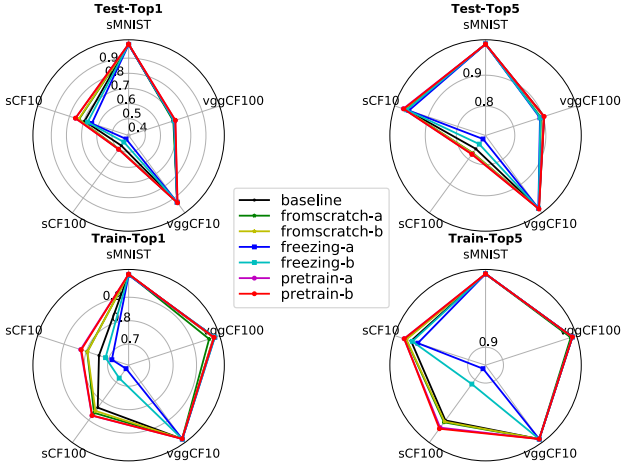

Fig. 7. **PS-Layer performance summary**. Classification accuracy overview of each baseline CNN-model *versus* our PS-Layer versions. Each circle summarizes the metrics performance (*Test*-Top1, *Test*-Top5, *Train*-Top1, *Train*-Top5) of each case study analyzed. Accuracy values were normalized between [0-1].

Layer *pre-train* versions (in magenta and red) outperforms baseline CNN-model (black) in each case study analyzed. Our PS-Layer provides greater interpretive power to CNN models due to simplicity and clear geometric interpretation of the object typicality concept (see Fig. 2).

V. CONCLUSION

In this Ph.D. thesis we introduced and evaluated three models based on Prototype Theory foundations to propose semantic representations of object's categories and object's images: *i)* a Computational Prototype Model (CPM), *ii)* a novel Prototype-based Description Model (GSDP) and *iii)* and Prototype-based Classification Model. Experiments showed that our CPM model can capture the object's visual typicality and the central and peripheral meaning of objects' categories. Our novel GSDP[5] representation introduces a new approach to the semantic description of object's images; and experiments in large image dataset shows that it is discriminative, small dimensioned, and encodes the semantic information of category members. Our PS-Layer introduces the image typicality property in semantic category learning process and experiments conducted showed that it can outperform some CNN-Models architecture.

Acknowledgment: We would like to thank the PPGCC-UFMG, CAPES, CNPq and FAPEMIG for funding this work.

VI. PUBLICATIONS & AWARDS

Parts of this work were published on the IEEE Winter Conference on Applications of Computer Vision (WACV) 2019, and an journal paper is under review on the Transactions on Image Processing (TIP). The thesis related to this work was also awarded to be presented at the WACV 2019 Doctoral Consortium.

[5]All source code, prototypes datasets, and tutorials are publicly available in our lab's Github: https://github.com/verlab/gsdp.

REFERENCES

[1] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346–359, 2008.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004.

[4] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2008, pp. 1–8.

[5] E. Tulving, "Coding and representation: searching for a home in the brain," *Science of Memory: Concepts*, pp. 65–68, 2007.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[8] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning." in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 4278–4284.

[9] V. Fromkin, R. Rodman, and N. Hyams, *An introduction to language*. Cengage Learning, 2018.

[10] E. Rosch, "Cognitive representations of semantic categories." *Journal of Experimental Psychology: General*, vol. 104, no. 3, p. 192, 1975.

[11] ——, "Principles of categorization," in *Cognition and Categorization*, E. Rosch and B. B. Lloyd, Eds. Hillsdale, NJ:Lawrence Erlbaum Associates, 1978, pp. 27– 48.

[12] E. Rosch and C. B. Mervis, "Family resemblances: Studies in the internal structure of categories," *Cognitive psychology*, vol. 7, no. 4, pp. 573–605, 1975.

[13] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4353–4361.

[14] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "Lift: Learned invariant feature transform," in *Proceedings of the of the European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 467–483.

[15] K. Lin, J. Lu, C.-S. Chen, and J. Zhou, "Learning compact binary descriptors with unsupervised deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1183–1192.

[16] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, and F. Moreno-Noguer, "Discriminative learning of deep convolutional feature point descriptors," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 118–126.

[17] S. Kim, D. Min, B. Ham, S. Lin, and K. Sohn, "Fcss: Fully convolutional self-similarity for dense semantic correspondence," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2018.

[18] I. Rocco, R. Arandjelović, and J. Sivic, "End-to-end weakly-supervised semantic alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[19] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 33, no. 5, pp. 978–994, 2011.

[20] E. Rosch, "Coherences and categorization: A historical view," *The development of language and language researchers: Essays in honor of Roger Brown*, pp. 373–392, 1988.

[21] D. Geeraerts, *Theories of lexical semantics*. Oxford University Press, 2010.

[22] S. R. Zaki, R. M. Nosofsky, R. D. Stanton, and A. L. Cohen, "Prototype and exemplar accounts of category learning and attentional allocation: A reassessment," *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 29, no. 6, pp. 1160–1173, 2003.

[23] E. Rosch and B. B. Lloyd, *Cognition and categorization*. Lawrence Erlbaum Associates Hillsdale, NJ, 1978, vol. 1.

[24] T. Kohonen, "Self-organization and associative memory," *Springer-Verlag Berlin Heidelberg New York. Also Springer Series in Information Sciences*, vol. 8, 1988.

[25] S. Seo and K. Obermayer, "Soft learning vector quantization," *Neural computation*, vol. 15, no. 7, pp. 1589–1604, 2003.

[26] P. Wohlhart, M. Köstinger, M. Donoser, P. M. Roth, and H. Bischof, "Optimizing 1-nearest prototype classifiers," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013, pp. 460–467.

[27] S. Jetley, B. Romera-Paredes, S. Jayasumana, and P. Torr, "Prototypical priors: From improving classification to zero-shot learning," in *Proceedings of the of the British Machine Vision Conference (BMVC)*, 2015.

[28] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, 2017, pp. 4080–4090.

[29] D. L. Medin and M. M. Schaffer, "Context theory of classification learning." *Psychological review*, vol. 85, no. 3, p. 207, 1978.

[30] J. P. Minda and J. D. Smith, "Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 28, no. 2, p. 275, 2002.

[31] B. Stellato, B. P. Van Parys, and P. J. Goulart, "Multivariate chebyshev inequality with estimated mean and variance," *The American Statistician*, vol. 71, no. 2, pp. 123–127, 2017.

[32] A. Martin, "The representation of object concepts in the brain," *Annual Review of Psychology*, vol. 58, pp. 25–45, 2007.

[33] J. A. Collins and K. M. Curby, "Conceptual knowledge attenuates viewpoint dependency in visual object recognition," *Visual Cognition*, vol. 21, no. 8, pp. 945–960, 2013.

[34] O. Pino, E. Nascimento, and M. Campos, "Prototypicality effects in global semantic description of objects," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan 2019, pp. 1233–1242.

[35] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.

[36] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on cifar-10," *Unpublished manuscript*, vol. 40, 2010.

[37] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[38] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[39] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on*. IEEE, 2015, pp. 730–734.

[40] B. Lake, W. Zaremba, R. Fergus, and T. Gureckis, "Deep neural networks predict category typicality ratings for images," in *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, 2015.

[41] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision (IJCV)*, vol. 42, no. 3, pp. 145–175, 2001.

[42] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 24, no. 7, pp. 971–987, 2002.

[43] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. IEEE, 2005, pp. 886–893.

[44] M. Li, "Texture moment for content-based image retrieval," in *2007 IEEE International Conference on Multimedia and Expo*. IEEE, 2007, pp. 508–511.

[45] Y.-j. Song, W.-b. Park, D.-w. Kim, and J.-h. Ahn, "Content-based image retrieval using new color histogram," in *Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004. Proceedings of 2004 International Symposium on*. IEEE, 2004, pp. 609–611.

[46] R. M. Haralick, K. Shanmugam *et al.*, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. 6, no. 6, pp. 610–621, 1973.

[47] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.