

Aprendizaje de una sola vez

El **aprendizaje de una sola vez** es un problema de categorización de objetos , que se encuentra principalmente en la visión por computadora . Mientras que la mayoría de los algoritmos de categorización de objetos basados en el aprendizaje automático requieren capacitación en cientos o miles de ejemplos, el aprendizaje único tiene como objetivo clasificar objetos a partir de uno o solo unos pocos ejemplos. El término **aprendizaje de pocos disparos** también se usa para estos problemas, especialmente cuando se necesita más de un ejemplo.

Motivación

La capacidad de aprender categorías de objetos a partir de pocos ejemplos y a un ritmo rápido se ha demostrado en humanos. ^[1] ^[2] Se estima que un niño aprende casi todas las 10 a 30 mil categorías de objetos del mundo a los seis años. ^[3] Esto se debe no solo al poder computacional de la mente humana, sino también a su capacidad para sintetizar y aprender nuevas categorías de objetos a partir de la información existente sobre diferentes categorías previamente aprendidas. Dados dos ejemplos de dos categorías de objetos: uno, un objeto desconocido compuesto de formas familiares, el segundo, una forma amorfa desconocida; es mucho más fácil para los humanos reconocer lo primero que lo segundo, lo que sugiere que los humanos hacen uso de categorías previamente aprendidas cuando aprenden otras nuevas. La motivación clave para resolver el aprendizaje de una sola vez es que los sistemas, como los humanos, pueden usar el conocimiento sobre categorías de objetos para clasificar nuevos objetos. ^[4] ^[5]

Antecedentes

Al igual que con la mayoría de los esquemas de clasificación , el aprendizaje de una sola vez implica tres desafíos principales:

- Representación: ¿Cómo se deben describir los objetos y las categorías?
- Aprendizaje: ¿Cómo se pueden crear tales descripciones?
- Reconocimiento: ¿Cómo se puede filtrar un objeto conocido del desorden envolvente, independientemente de la oclusión, el punto de vista y la iluminación? ^[6]

El aprendizaje de una sola vez difiere del reconocimiento de un solo objeto y de los algoritmos de reconocimiento de categorías estándar en su énfasis en la transferencia de conocimiento, que hace uso de categorías aprendidas previamente.

- Parámetros del modelo: reutiliza los parámetros del modelo, en función de la similitud entre las categorías antiguas y nuevas. Las categorías se aprenden primero en numerosos ejemplos de entrenamiento, luego se aprenden nuevas categorías usando transformaciones de los parámetros del modelo de esas categorías iniciales o seleccionando parámetros relevantes para un clasificador. ^[7]
- Uso compartido de funciones: comparte partes o funciones de objetos en todas las categorías. Un algoritmo extrae "información de diagnóstico" en parches de categorías ya aprendidas maximizando la información mutua de los parches y luego aplica estas características al aprendizaje de una nueva categoría. Una categoría de perro, por ejemplo, se puede aprender de

una sola vez a partir del conocimiento previo de las categorías de caballo y vaca, porque los objetos de perro pueden contener parches distintivos similares. ^[8]

- Información contextual: Apela al conocimiento global de la escena en la que aparece el objeto. Dicha información global se puede utilizar como distribuciones de frecuencia en un marco de campo aleatorio condicional para reconocer objetos. ^[9] Alternativamente, el contexto puede considerar la altura de la cámara y la geometría de la escena. ^[10] Los algoritmos de este tipo tienen dos ventajas. Primero, aprenden categorías de objetos que son relativamente diferentes; y segundo, funcionan bien en situaciones ad hoc donde una imagen no ha sido recortada y alineada a mano. ^[11]

Teoría

El algoritmo de aprendizaje one-shot bayesiano representa el primer plano y el fondo de las imágenes parametrizadas por una combinación de modelos de constelaciones. ^[12] Durante la fase de aprendizaje, los parámetros de estos modelos se aprenden utilizando un parámetro de densidad conjugado posterior y la maximización de la expectativa bayesiana variacional (VBEM). ^[13] En esta etapa, las categorías de objetos previamente aprendidas informan la elección de los parámetros del modelo a través de la transferencia de información contextual. Para el reconocimiento de objetos en nuevas imágenes, la posterior obtenida durante la fase de aprendizaje se utiliza en un marco de decisión bayesiano para estimar la relación de $p(\text{objeto} \mid \text{probar, entrenar})$ $ap(\text{desorden de fondo} \mid \text{probar, entrenar})$ donde p es la probabilidad del resultado. ^[14]

Marco bayesiano

Dada la tarea de encontrar un objeto particular en una imagen de consulta, el objetivo general del algoritmo de aprendizaje one-shot bayesiano es comparar la probabilidad de que el objeto esté presente frente a la probabilidad de que solo esté presente el desorden de fondo. Si la primera probabilidad es mayor, el algoritmo informa la presencia del objeto, de lo contrario, el algoritmo informa su ausencia. Para calcular estas probabilidades, la clase de objeto debe modelarse a partir de un conjunto de (1 ~ 5) imágenes de entrenamiento que contienen ejemplos.

Para formalizar estas ideas, dejemos I ser la imagen de consulta, que contiene un ejemplo de la categoría de primer plano O_{fg} o solo desorden de fondo de una categoría de fondo genérica O_{bg} . también deja I_t ser el conjunto de imágenes de entrenamiento utilizadas como categoría de primer plano. La decisión de si I contiene un objeto de la categoría de primer plano, o solo el desorden de la categoría de fondo es:

$$R = \frac{p(O_{fg} \mid I, I_t)}{p(O_{bg} \mid I, I_t)} = \frac{p(I \mid I_t, O_{fg})p(O_{fg})}{p(I \mid I_t, O_{bg})p(O_{bg})},$$

donde la clase posterior $p(O_{fg} \mid I, I_t)$ y $p(O_{bg} \mid I, I_t)$ han sido expandidos por el teorema de Bayes, produciendo una relación de probabilidades y una relación de categorías de objetos previas. Decidimos que la imagen I contiene un objeto de la clase de primer plano si R supera un cierto umbral T . A continuación, presentamos modelos paramétricos para las categorías de primer plano y segundo plano con parámetros θ y θ_{bg} respectivamente. Este modelo paramétrico de primer plano se aprende

durante la etapa de aprendizaje de I_t , así como información previa de las categorías aprendidas. El modelo de fondo que asumimos es uniforme en todas las imágenes. Omitiendo la proporción constante de categorías previas, $\frac{p(O_{fg})}{p(O_{bg})}$, y parametrizando sobre θ y θ_{bg} rendimientos

$$R \propto \frac{\int p(I|\theta, O_{fg})p(\theta|I_t, O_{fg})d\theta}{\int p(I|\theta_{bg}, O_{bg})p(\theta_{bg}|I_t, O_{bg})d\theta_{bg}} = \frac{\int p(I|\theta)p(\theta|I_t, O_{fg})d\theta}{\int p(I|\theta_{bg})p(\theta_{bg}|I_t, O_{bg})d\theta_{bg}}, \text{ habiendo simplificado } p(I|\theta, O_{fg})yp(I|\theta, O_{bg})ap(I|\theta_{fg})yp(I|\theta_{bg}).$$

La distribución posterior de los parámetros del modelo dadas las imágenes de entrenamiento, $p(\theta|I_t, O_{fg})$ se estima en la fase de aprendizaje. En esta estimación, el aprendizaje de un paso difiere marcadamente de los modelos de estimación bayesianos más tradicionales que aproximan la integral como $\delta(\theta^{ML})$. En su lugar, utiliza un enfoque variacional utilizando información previa de categorías previamente aprendidas. Sin embargo, la estimación tradicional de máxima verosimilitud de los parámetros del modelo se utiliza para el modelo de fondo y las categorías aprendidas de antemano a través del entrenamiento. [15]

Modelo de categoría de objetos

Para cada imagen de consulta I e imágenes de entrenamiento I_t , se utiliza un modelo de constelación para la representación. [12] [16] [17] Para obtener este modelo para una imagen dada I , primero se detecta un conjunto de N regiones interesantes en la imagen usando el detector de prominencia de bradi de Kadir. [18] Cada región seleccionada está representada por una ubicación en la imagen, X_i y

una descripción de su apariencia, A_i . Alquiler $X = \sum_{i=1}^N X_i$, $A = \sum_{i=1}^N A_i$ y X_t y A_t las representaciones análogas para imágenes de entrenamiento, la expresión para R se convierte en:

$$R \propto \frac{\int p(X, A|\theta, O_{fg})p(\theta|X_t, A_t, O_{fg})d\theta}{\int p(X, A|\theta_{bg}, O_{bg})p(\theta_{bg}|X_t, A_t, O_{bg})d\theta_{bg}} = \frac{\int p(X, A|\theta)p(\theta|X_t, A_t, O_{fg})d\theta}{\int p(X, A|\theta_{bg})p(\theta_{bg}|X_t, A_t, O_{bg})d\theta_{bg}}$$

las probabilidades $p(X, A|\theta)$ y $p(X, A|\theta_{bg})$ se representan como mezclas de modelos de constelaciones. Un modelo de constelación típico tiene $P(3 \sim 7)$ partes, con $N(\sim 100)$ regiones de interés. **Por lo tanto, un vector \mathbf{h}** P -dimensional asigna una región de interés (de N regiones) a cada parte del modelo (para P partes). Por lo tanto, \mathbf{h} denota una hipótesis (una asignación de regiones de interés a las partes del modelo) para el modelo y un modelo de constelación completa se representa sumando todas las hipótesis posibles \mathbf{h} en el espacio de hipótesis H . Finalmente la probabilidad se escribe

$$p(X, A|\theta) = \sum_{\omega=1}^{\Omega} \sum_{\mathbf{h} \in H} p(X, A, \mathbf{h}, \omega|\theta).$$

Lo diferentes ω 's representan diferentes configuraciones de partes, mientras que las diferentes hipótesis \mathbf{h} representan diferentes asignaciones de regiones a partes, dado un modelo de parte ω . La suposición de que la forma del modelo (representada por \mathbf{X} , la colección de ubicaciones de partes) y la apariencia son independientes permite considerar la expresión de probabilidad $p(\mathbf{X}, \mathbf{A}, \mathbf{h}, \omega | \theta)$ como dos probabilidades separadas de apariencia y forma. ^[19]

Apariencia

La apariencia de cada característica está representada por un punto en el espacio de apariencia (discutido a continuación en la implementación). "Cada parte p en el modelo de constelación tiene una densidad gaussiana dentro de este espacio con parámetros medios y de precisión $\theta_{p,\omega}^A = \mu_{p,\omega}^A, \Gamma_{p,\omega}^A$." A partir de estos, la probabilidad de aparición descrita anteriormente se calcula como un producto de gaussianos sobre las partes del modelo para una hipótesis dada \mathbf{h} y un componente de mezcla ω . ^[20]

Shape

The shape of the model for a given mixture component ω and hypothesis \mathbf{h} is represented as a joint Gaussian density of the locations of features. These features are transformed into a scale and translation-invariant space before modelling the relative location of the parts by a $2(P - 1)$ -dimensional Gaussian. From this, we obtain the shape likelihood, completing our representation of $p(\mathbf{X}, \mathbf{A}, \mathbf{h}, \omega | \theta)$. In order to reduce the number of hypotheses in the hypothesis space \mathbf{H} , only those hypotheses that satisfy the ordering constraint that the x-coordinate of each part is monotonically increasing are considered. This eliminates $P!$ hypotheses from \mathbf{H} . ^[20]

Conjugate densities

In order to compute R , the integral $\int p(\mathbf{X}, \mathbf{A} | \theta) p(\theta | \mathbf{X}_t, \mathbf{A}_t, \mathbf{O}_{fg}) d\theta$ must be evaluated, but is analytically intractable. The object category model above gives information about $p(\mathbf{X}, \mathbf{A} | \theta)$, so what remains is to examine $p(\theta | \mathbf{X}_t, \mathbf{A}_t, \mathbf{O})$, the posterior of θ , and find a sufficient approximation to render the integral tractable. Previous work approximates the posterior by a δ function centered at θ^* , collapsing the integral in question into $p(\mathbf{X}, \mathbf{A} | \theta^*)$. This θ^* is normally estimated using a Maximum Likelihood ($\theta^* = \theta^{ML}$) or Maximum A Posteriori ($\theta^* = \theta^{MAP}$) procedure. However, because in one-shot learning, few training examples are used, the distribution will not be well-peaked, as is assumed in a δ function approximation. Thus instead of this traditional approximation, the Bayesian one-shot learning algorithm seeks to "find a parametric form of $p(\theta)$ such that the learning of $p(\theta | \mathbf{X}_t, \mathbf{A}_t, \mathbf{O}_{fg})$ is feasible". The algorithm employs a Normal-Wishart distribution as the conjugate prior of $p(\theta | \mathbf{X}_t, \mathbf{A}_t, \mathbf{O}_{fg})$, and in the learning phase, variational Bayesian methods with the same

computational complexity as maximum likelihood methods are used to learn the hyperparameters of the distribution. Then, since $p(\mathbf{X}, \mathbf{A}|\theta)$ is a product of Gaussians, as chosen in the object category model, the integral reduces to a multivariate Student's T distribution, which can be evaluated.^[21]

Implementation

Feature detection and representation

To detect features in an image so that it can be represented by a constellation model, the Kadir Brady feature detector is used on grey-scale images, finding salient regions of the image. These regions are then clustered, yielding a number of features (the clusters) and the shape parameter \mathbf{X} , composed of the cluster centers. The Kadir Brady detector was chosen because it produces fewer, more salient regions, as opposed to feature detectors like multiscale Harris, which produces numerous, less significant regions.

The regions are then taken from the image and rescaled to a small patch of 11 x 11 pixels, allowing each patch to be represented in 121-dimensional space. This dimensionality is reduced using principal component analysis, and \mathbf{A} , the appearance parameter, is then formed from the first 10 principal components of each patch.^[22]

Learning

To obtain shape and appearance priors, three categories (spotted cats, faces, and airplanes) are learned using maximum likelihood estimation. These object category model parameters are then used to estimate the hyper-parameters of the desired priors.

Given a set of training examples, the algorithm runs the feature detector on these images, and determines model parameters from the salient regions. The hypothesis index \mathbf{h} assigning features to parts prevents a closed-form solution of the linear model, so the posterior $p(\theta|\mathbf{X}_t, \mathbf{A}_t, \mathbf{O}_{fg})$ is estimated by variational Bayesian expectation-maximization, which is run until parameter convergence after ~ 100 iterations. Learning a category in this fashion takes under a minute on a 2.8 GHz machine with a 4-part model and < 10 training images.^[23]

Experimental results

Motorbike example

To learn the motorbike category:

- Six training images are selected from the motorbike category of the Caltech 4 Data Set and the Kadir Brady detector is applied, giving \mathbf{X}_t and through PCA, \mathbf{A}_t .
- Next, the prior model parameters are computed from 30 models θ_t , 10 from each of the three learned categories: spotted cats, faces, and airplanes. This prior encodes the knowledge that "models lacking visual consistency [ie background clutter] occupy a different part of the parameter space [from] coherent models."

- In learning, which is performed next, the prior biases the posterior $p(\theta|X_t, A_t, O_{fg})$ towards parts of the parameter space corresponding to coherent models. Only one mixture component is used, letting $\Omega = \mathbf{1}$. The estimation of the posterior is shown below.
- Finally, the figures below show the learned motorbike model with shape and appearance of parts, and the corresponding features.
- For recognition tests, the model above is applied to 50 images which contain motorbikes and 50 which do not. The image below shows an ROC curve, measuring the probability of detection over the probability of false detection, as well as some recognized examples.

Shared densities on transforms

Another algorithm uses knowledge transfer by model parameters to learn a new object category that is similar in appearance to previously learned categories. An image is represented as either a texture and shape, or as a latent image that has been transformed, denoted by $I = T(I_L)$.

A Siamese neural network works in tandem on two different input vectors to compute comparable output vectors.^[24]

Congealing

In this context, congealing is "the simultaneous vectorization of each of a set of images to each other". For a set of training images of a certain category, congealing iteratively transforms each image to minimize the images' joint pixelwise entropies E , where

$$E = \sum_{p=1}^P H(\nu(p)),$$

"where $\nu(p)$ is the binary random variable defined by the values of a particular pixel p across all of the images, $H()$ is the discrete entropy function of that variable, and $\mathbf{1} \leq p \leq P$ is the set of pixel indices for the image."

The congealing algorithm begins with a set of images I_i and a corresponding transform matrix U_i , which at the end of the algorithm will represent the transformation of I_i into its latent I_{L_i} . These latents I_{L_i} minimize the joint pixel-wise entropies. Thus the task of the congealing algorithm is to estimate the transformations U_i .

Sketch of algorithm:

- Initialize U_I 's to the identity.
- Compute the joint pixelwise entropies of the current set of images.
- For each image I_i , iterate through all possible affine transformations A (rotation, x-translation, y-translation, x-scale, y-scale, x-shear, y-shear) and test if AU_i decreases the joint pixelwise entropies. If so, set $U_i = AU_i$.
- Repeat previous step until convergence.

At the end of the algorithm, $U_i(I) = I_{L_i}$, and $T = U_i^{-1}$ transforms the latent image back into the originally observed image.^[25]

Classification

To use this model for classification, it must be estimated with the maximum posterior probability given an observed image I . Applying Bayes' rule to $P(c_j|I)$ and parametrization by the transformation T gives a difficult integral that must be approximated, and then the best transform T (that which maps the test image to its latent image) must be found. Once this transformation is found, the test image can be transformed into its latent, and a nearest neighbor classifier based on Hausdorff distance between images can classify the latent (and thus the test image) as belonging to a particular class c_j .

To find T , the test image I is inserted into the training ensemble for the congealing process. Since the test image is drawn from one of the categories c_j , congealing provides a corresponding $T_{\text{test}} = U_{\text{test}}^{-1}$ that maps I to its latent. The latent can then be classified.^[26]

Single-example classification

Given a set of transformations B_i obtained from congealing many images of a certain category, the classifier can be extended to the case where only one training I_t example of a new category c is allowed. Applying all the transformations B_i sequentially to I_t creates an artificial training set for c . This artificial data set can be made larger by borrowing transformations from many already known categories. Once this data set is obtained, I , a test instance of c , can be classified as in the normal classification procedure. The key assumption is that categories are similar enough that the transforms from one can be applied to another.^[27]

In natural language processing

With the introduction of the practice of scaling up language models like OpenAI's GPT-2 and GPT-3, few-shot performance of such models has been shown to achieve competitive results on natural language processing (NLP) tasks, sometimes surpassing prior state-of-the-art fine-tuning approaches. Examples of such NLP tasks are translation, question answering, cloze tasks, unscrambling words, and using a novel word in a sentence.^[28] In these cases, one-shot and few-shot learning have been demonstrated purely via text interaction with the model. This is done by prepending examples directly to the prompt that is fed into the model for inference. The creation and optimisation of such prompts is called prompt engineering and is now an active field of study.

See also

- Variational Bayesian methods
- Variational message passing
- Expectation-maximization algorithm
- Bayesian inference
- Feature detection
- Association rule learning

- Hopfield network
- Zero-shot learning

Citations

1. Li, Fergus & Perona 2002.
2. Thorpe, Fize & Marlot 1996.
3. Biederman 1987.
4. Li, Fergus & Perona 2006, Section 1.
5. Li 2006, Section 1.
6. Li, Fergus & Perona 2006, Section 2.
7. Fink 2004.
8. Bart & Ullman 2005.
9. Murphy & et al 2004.
10. Hoiem, Efros & Herbert 2005.
11. Li 2006, Section 2.
12. Burl & et al 1996.
13. Attias 1999.
14. Li & et al 2006.
15. Li, Fergus & Perona 2006, Section 3.1.
16. Weber, Welling & Perona 2000.
17. Fergus, Perona & Zisserman 2003.
18. Kadir & Brady 2001.
19. Li, Fergus & Perona 2006, Section 3.2.
20. Li, Fergus & Perona 2006, Section 3.2.1.
21. Li, Fergus & Perona 2006, Section 3.4.3.
22. Li, Fergus & Perona 2006, Section 5.1.
23. Li, Fergus & Perona 2006, Sections 4, 5.2.
24. *Few-Shot Learning (2/3): Siamese Networks* (<https://www.youtube.com/watch?v=4S-XDefSjTM>). *YouTube*. Archived (<https://ghostarchive.org/varchive/youtube/20211210/4S-XDefSjTM>) from the original on 2021-12-10.
25. Miller et al.
26. Miller, Matsakis & Viola 2000, Section 4.
27. Miller, Matsakis & Viola 2000, Section 7.
28. Brown, Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav; Sastry, Girish; Askell, Amanda; Agarwal, Sandhini; Herbert-Voss, Ariel; Krueger, Gretchen; Henighan, Tom; Child, Rewon; Ramesh, Aditya; Ziegler, Daniel M.; Wu, Jeffrey; Winter, Clemens; Hesse, Christopher; Chen, Mark; Sigler, Eric; Litwin, Mateusz; Gray, Scott; Chess, Benjamin; Clark, Jack; Berner, Christopher; McCandlish, Sam; Radford, Alec; Sutskever, Ilya; Amodei, Dario (22 July 2020). "Language Models are Few-Shot Learners". *arXiv:2005.14165* (<https://arxiv.org/abs/2005.14165>) [*cs.CL* (<https://arxiv.org/archive/cs>. *CL*)].

References

- Li, Fei Fei (2006). "Knowledge transfer in learning to recognize visual object classes" (http://vision.stanford.edu/documents/Li_ICDL2006.pdf) (PDF). *International Conference on Development and Learning (ICDL)*.
- Li, Fei Fei; Fergus, R.; Perona, P. (2006). "One-Shot learning of object categories" (<http://vision.stanford.edu/documents/LiFergusPerona2006.pdf>) (PDF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **28** (4): 594–611. doi:10.1109/TPAMI.2006.79 (<https://doi.org/10.1109/2FTPAMI.2006.79>). PMID 16566508 (<https://pubmed.ncbi.nlm.nih.gov/16566508>). S2CID 6953475 (<https://api.semanticscholar.org/CorpusID:6953475>).
- Miller; Matsakis; Viola (2000). "Learning from One Example through Shared Densities on Transforms" (<https://people.cs.umass.edu/~elm/papers/cvpr2000.pdf>) (PDF). *Proc. Computer Vision and Pattern Recognition*.
- Li, F.F.; VanRullen, R.; Coch, C.; Perona, P. (2002). "Rapid natural scene categorization in the near absence of attention" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC123186>). *PNAS*. **99** (14): 9596–9601. Bibcode:2002PNAS...99.9596L (<https://ui.adsabs.harvard.edu/abs/2002PNAS...99.9596L>). doi:10.1073/pnas.092277599 (<https://doi.org/10.1073%2Fpnas.092277599>). PMC 123186 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC123186>). PMID 12077298 (<https://pubmed.ncbi.nlm.nih.gov/12077298>).
- Thorpe, S.; Fize, D.; Marlot, C. (1996). "Speed of processing in the human visual system" (<https://www.allpsych.uni-giessen.de/rausch12/readings/ThorpeEtal.Nature.1996.pdf>) (PDF). *Nature*. **381** (6582): 520–522, 1996. Bibcode:1996Natur.381..520T (<https://ui.adsabs.harvard.edu/abs/1996Natur.381..520T>). doi:10.1038/381520a0 (<https://doi.org/10.1038%2F381520a0>). PMID 8632824 (<https://pubmed.ncbi.nlm.nih.gov/8632824>). S2CID 4303570 (<https://api.semanticscholar.org/CorpusID:4303570>).
- Biederman, I. (1987). "Recognition-by-Components: a theory of human understanding" (<http://courses.nus.edu.sg/course/elec/ee6901/Recognition%20by%20Components.pdf>) (PDF). *Psychological Review*. **94** (2): 115–147. doi:10.1037/0033-295X.94.2.115 (<https://doi.org/10.1037%2F0033-295X.94.2.115>). PMID 3575582 (<https://pubmed.ncbi.nlm.nih.gov/3575582>).
- Fink, M. (2004). "Object classification from a single example utilizing class relevance pseudo-metrics". *NIPS*. CiteSeerX 10.1.1.91.7461 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.91.7461>).
- Bart; Ullman (2005). "Cross-generalization: learning novel classes from a single example by feature replacement" (https://www.inf.tu-dresden.de/content/institutes/ki/is/HS_SS08_Papers/BartUllmanCVPR05.pdf) (PDF). *CVPR*.
- Murphy, K.; Torralba, A.; Freeman, W.T. (2004). "Using the forest to see the trees: a graphical model relating features, objects, and scenes" (https://www.cs.ubc.ca/~murphyk/Papers/nips2003_1ong.pdf) (PDF). *NIPS*.
- Hoiem, D.; Efros, A.A.; Herbert, M. (2005). "Geometric context from a single image" (https://www.cs.utexas.edu/~grauman/courses/spring2009/slides/luis_pres.pdf) (PDF). *ICCV*.
- Attias, H. (1999). "Inferring Parameters and Structure of Latent Variable Models by Variational Bayes" (<https://arxiv.org/pdf/1301.6676>). *Proc. Of the 15th Conf. In Uncertainty in Artificial Intelligence*: 21–30. arXiv:1301.6676 (<https://arxiv.org/abs/1301.6676>).
- Burl, M.; Weber, M.; Perona, P. (1996). "A Probabilistic Approach to Object Recognition Using Local Photometry and Global Geometry" (<https://link.springer.com/content/pdf/10.1007/BFb0054769.pdf>) (PDF). *Proc. European Conf. Computer Vision*. Lecture Notes in Computer Science. **1407**: 628–641. doi:10.1007/BFb0054769 (<https://doi.org/10.1007%2FBFb0054769>). ISBN 978-3-540-64613-6.
- Fergus, R.; Perona, P.; Zisserman, A. (2003). "Object Class Recognition by Unsupervised Scale-Invariant Learning" (<http://host.robots.ox.ac.uk/pascal/VOC/pubs/fergus03.pdf>) (PDF). *Proc. Computer Vision and Pattern Recognition*: 264–271.
- Weber, M.; Welling, M.; Perona, P. (2000). "Unsupervised Learning of Models for Recognition" (https://link.springer.com/content/pdf/10.1007/3-540-45054-8_2.pdf) (PDF). *Proc. European Conf.*

Computer Vision. Lecture Notes in Computer Science. **1842**: 101–108. doi:10.1007/3-540-45054-8_2 (https://doi.org/10.1007%2F3-540-45054-8_2). ISBN 978-3-540-67685-0.

- Kadir, T.; Brady, M. (2001). "Scale, Saliency, and Image Description" (<https://www.researchgate.net/publication/220660282>). *International Journal of Computer Vision*. **45** (2): 83–105. doi:10.1023/A:1012460413855 (<https://doi.org/10.1023%2FA%3A1012460413855>). S2CID 825395 (<https://api.semanticscholar.org/CorpusID:825395>).
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=One-shot_learning&oldid=1129501511"