Methods of Pornography Detection: Review

2 authors:

Sasan Karamizadeh
Iran Telecommunication Research Center
30 PUBLICATIONS **431** CITATIONS

SEE PROFILE

Abouzar Arabsorkhi
Iran Telecommunication Research Center
13 PUBLICATIONS **26** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project — Providing industrial decision making system to prioritize the MTO and MTS orders View project

Project — Work Projects View project

Methods of Pornography Detection: Review

SASAN KARAMIZADEH
Iran Telecommunication Research Center,
Tehran, Iran
s.karamizadeh@itrc.ac.ir

ABOUZAR ARABSORKHI
Faculty member of Iran Telecommunication
Research Center, Tehran, Iran
abouzar_arab@itrc.ac.ir

ABSTRACT
In recent years, prone images and other such indecent matter are available on the social media and the Internet for children. Filtering of image porn has become one of the big changes for searches; they are tied to finding methods to filter porn images. Social media network is interested in filter porn images from normal ones. Analysis method uses the bright image to automatically detect and filter images in the media. In this paper, we have reviewed methods such as color based, shape based, local and global feature approach, deep learning and bag-of-words for filtering porn images which include comparing with the advantages and disadvantages.

CCS Concepts
• **Security and privacy**→**Social aspects of security and privacy.**

Keywords
Internet; filtering; porn image; deep learning; color based.

1. INTRODUCTION
Filtering complex media (violent, pornographic, etc.) has become significant because of the intense utilization of online media by individuals of any age; and among delicate media sorts, pornography entertainment is regularly the most unwelcome. A kind of utilizations has expanded societal enthusiasm on the issue, e.g., recognizing unseemly conduct by means of observation cameras; or diminishing the trading of sexually-charged texts, otherwise called "sexting", by minors [1]. Furthermore, law implementers may utilize explicit entertainment channels as a first filter when searching for teenager pornography entertainment in the criminological examination of computers or internet content. There are two major methods of identifying images related to porn namely i) concentrating on web page contents with that picture that will arrange the picture content alternately; ii) they search inside the images [2]. They judge these contents using skin texture pixels [3]. In this paper, we reviewed several pornography filtering detection methods.

2. COLOR BASED
The color-based approach depends on the activity suspicion that pixels in pornographic images are mostly skin [4]. Color models are normally used to characterize a vigorous skin color

representation [5].Skin proportion, histogram, color probabilities, and associated segments are utilized to describe skin color. The most vital feature used to control an image that contains porn or otherwise is color [6]. Skin filter is one of the important parts of a porn detection system, which expects to determine which pixels are skin colored or not utilizing a skin color appropriation model [7]. To classify the skin pixels as or otherwise not of skin color, the models on the distribution of skin utilize a space with a single color [8, 9]. This is built on the result to show the selection of space's color is insignificant if a thresholding is carried using enough training as well as a proper false alarm. It can be divided into three subcategories: color range computed color histograms, and parametric color distribution functions [10, 11].

2.1 Computed Color Histograms
There are two problems that must be tackled in order to build a histogram model: the choice of the size of the histogram and color space, which is measured by the number of channels in the color channel [11]. Color histogram is verified for small databases. Only color information is recorded by a color histogram; images can be completely different with similar color histograms [12]. Minor computing and robustly tolerant movement of an object in an image are advantages of the color histograms [11]. A color histogram is a vector where every passage stores the quantity of pixels of a given color in the image [12]. Histogramming is applied after all the images have been scaled to contain a similar number of pixels. Colors are mapped into a discrete color spade space containing colors [13, 14].

3. SHAPE BASED
One of the important characteristics of an object is shape. Shapes have been extracted from skin regions in the shape-based method. The exclusively describe the object shape is gold of shape descriptor. Shape descriptor has to minimize the maximize the between-class variance and be insensitive to noise [15]. Boundary, region and moment have represented in shape [15]. For making measurements of shapes, or recognition objects and matching shapes can be utilized these representations. The contour-based feature, Geometric constraints, Gaussian model and Hu and Zernike moments of the skin distribution are four categories of shape based [16, 17]; Figure 1 shows shape based on image.

(a)   (b)

Figure 1. skin region detection. (a). Original image. (b) Binary skin image.

3.1 Contour-Based Features

Boundary information is extracted by using contour based features techniques. The contour based shape can be separated into structure and global methods [18]. The global method does not divide the shape into subparts to drive the feature vector; a matching process is used to complete the boundary information, therefore known as the continuous approach [19]. The shape is broken down using a structural approach into subparts, and this approach is known as the separate method. Normally, the structural approach is represented by a string or tree and graph, which is finally used to reconcile the image retrieval process [20].

4. LOCAL AND GLOBAL FEATURE APPROACH

Essentially, the local and global features contain two types of feature extractions from the image based on application. Usually, image retrieval is utilized in global descriptors, classification, and object detection, whereas local is used for the object [21, 22]. For classifying normal and pornographic images, a local description using a local feature approach is applied [9].

Whole image is described by global features to generalize the complete object while image patches are used in local features for the description of an object. Texture in an image patch is represented by the local features. Contour representation, texture feature, and shape descriptors are part of the global features. The feature method uses the Scale Invariant Feature Transform (SIFT) as well as the Probabilistic Latent Semantic Analysis (PLSA) [22-24]. Figure 2 shows global and local image descriptors.
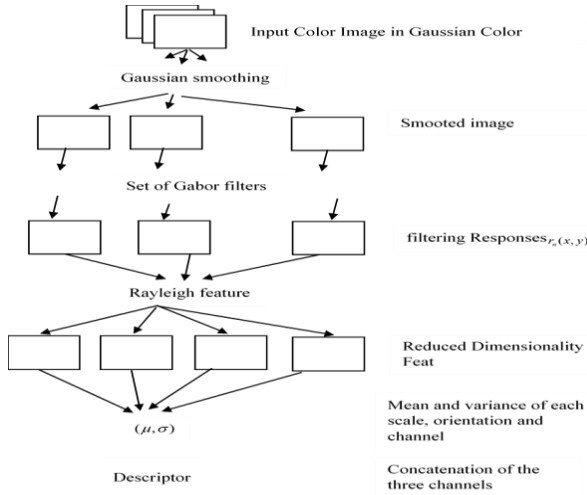

Figure 2. Color Texture Description scheme.

4.1 Probabilistic Latent Semantic Analysis

Stops from a statistical opinion of LSA is opposed to the standard PLSA; the great generative data model is defined by PLSA [25]. PLSA has a lot of advantages: On the most general level, it infers that standard methods from statistics can be connected for model fitting, difficulty control, and model selection [26]. For instance, one can evaluate the value of a PLSA model by measuring its foreboding performance, e.g., with the assistance of cross-validation. Figure 3 shows The general structure of pLSA model. This shows the intermediate layer of latent topics that links the documents and the words: each document can be represented as a mixture of concepts weighted by the probability P(zjd) and each word expresses a topic with probability P(wjz) [27]. Likely pLSA's structure is statistical model which is called aspect model.
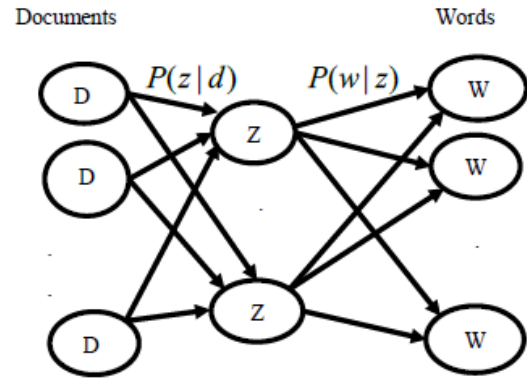

Figure 3. The general structure of pLSA model.

4.2 Scale Invariant Feature Transform

First calculation of the orientation of the most dominant gradient is used for computing the SIFT feature [28]. At that point, in respect to this coordination, the feature vector passages based on the gradient are developed from the gray scaled neighborhood that is local by dividing the local neighborhood into sub-regions of 4 × 4, subsequently gathering each pixel's gradient magnitude into a histogram with local orientation [29]. The Gaussian window is used to weigh the gradients concentrating on the location of the intriguing point. The SIFT algorithm is used for the image feature generation, which is partly invariant to affine projection and illumination changes and invariant to scaling, rotation, image translation [30]. SIFT can be divided into four steps: key point localization, key point descriptor, orientation assignment, and scale-space extrema detection. The advantages of SIFT are 1. It is routed to clutter and occlusion because the Features are local; 2. In the large database of objects, Individual features can be matched; 3. Small objects can be used to create many features; 4. Performance is near to real-time; and 5. With each added robustness, an extensive range of different feature types can easily be extracted [31]. Figure 4 shows of how the SIFT descriptor is computed from sampled values of the gradient orientation and the gradient magnitude over a locally adapted grid around each interest point, with the scale factor determined from the detection scales of the interest point and the orientation determined from the dominant peak in a gradient orientation histogram around the interest point. This figure illustrations an image descriptor computed over a 2×2 whereas the SIFT descriptor is usually computed over a 4×4 grid [28].
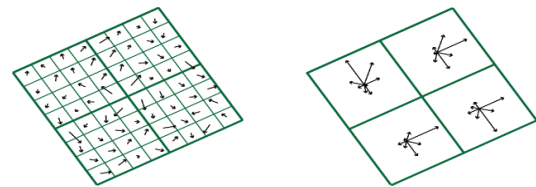

Figure 4. SIFT descriptor.

5. DEEP LEARNING

Computational model is used by deep learning to compose multiprocessing layers to learn representations of data with multiple levels of abstraction [32].

Most neural networks exist due to the availability of data and the cost of computation is a single layer. To solve pattern recognition

problems, the deep learning technique is utilized, which uses multiple layers for learning in the neural networks. Deep learning can be divided into four categories. Multiple levels of representation are utilized by the deep-learning method to represent the learning method [33].

5.1 CNN-Based Methods
One of the famous deep learning methods is the Convolutional Neural Networks where multiple layers are trained in a strong manner. This method is utilized by under the computer vision because it has been greatly effective [34]. The Convolutional layer is used by several kernels to convolve the complete image in addition to the intermediate creating numerous feature maps; pooling layer follows a convolutional layer for reducing overfitting and a fully-connected layer flow, finally the network layers for classification [35, 36]. CNN includes five sub-categories which are The adopted Alexnet-based[37] which is one of the important CNN's architecture; GoogleNet is increased depth and width without increasing calculations Requirements [38]; SPP are attached layers are fully suggested to "collapse the space pyramid" to remove the need for the image Clarity; VGG which is for checking an accurate network of depth-increasing grids [39], Clarifai is used for insight into the function of intermediate feature layers[40].Table1 shows the configuration of CNN sub-categories

Table 1. shows the configurations of CNN sub-categories

VGG	15 convolutional layers and 3 totally connected layers [36]
SPP	5 layers that are convolutional and 3 totally connected layers [35]
Alexnet-based	5 layers that are convolutional and wholly connected layers [33]
Google Net	21 layers that are convolutional and 1 totally connected layer[38]
Clarifai	5 convolutional layers and 3 fully connected layers [37]

5.2 Auto Encoding-Based Methods
Different kinds of artificial neural networks are utilized to study effective encoding, which includes the auto-encoder [41]. Substitute for preparing the system to foresee some objective esteem Y given sources of info X, an auto-encoder is prepared to remake the inputs itself. Thus, the output vectors contain similar dimensions as the input vector. Normally, a single layer has been unable to receive representative feature and act discriminatively from the raw data [42]. An Auto encoding Deep auto-encoder is utilized; this then pushes forward the code, which sends the learned code from the past auto-encoder to the following one to achieve its tasks. Sparse auto-encoding is one of supervised methods to learn feature from the unlabeled data automatically [43], de noising auto encoding makes the recovery with correct input from the ruined sort [44]; and contractive auto encoder: robustness is achieved through the addition of the penalty on analytic contractions to the reconstruction error's function. [45] are the sub-categories of auto encoding.

5.3 Sparse Coding-Based Methods
One of the unsupervised methods for learning sets of over complete basis to represent data comfortability is sparse coding.

The sparse coding aims to uncover a basic vector set ϕi which is

able to represent the input vector x, as these linear combination's basis vectors [46].

$$x = \sum\nolimits_{}^{} i = 1 k a i\, \phi i$$

Basic vectors are able to capture inherent patterns and structures in the input data by having an over-complete basic. Sparse coding includes i) sparse coding SPM, which has been widely applied in the processing of sensory data, such as acoustic and image signals [47]; ii) local coordinate coding - The main idea is to locally embed points on the manifold into a lower dimensional space, expressed as coordinates with respect to a set of anchor points [48]; iii) super-vector coding which is algorithmically a simple extension of Vector Quantization (VQ) coding [49]; and iv) laplacian sparse coding where the same features are not just provided for the assigned to cluster centers that are optimally-selected, which again guarantees the chosen cluster centers to be the same [50, 51].

5.4 RBM-Based Methods
Single layer of hidden units is an RBM with no connection to each other with undirected, regular connections to a layer of visible units. Created date is stated from the RBM using a starting random state in one of the layers and then performing an alternating Gibbs sampling [52]. All units in one have been updated with parallel given to the present states of the units in the other layer. This will be repeated until the system samples from its distribution equilibrium. This can be divided into category deep belief networks, which includes a model that is probabilistic and generative that offer a joint distribution of probability over the absorbable labels as well as data [53], deep Boltzmann machines include another learning algorithm where the units are arranged in layers [54], and the mode, swotj with energy models are deep with just a single layer of stochastic hidden units to perform efficient training as well as inferences [55].

Table 2. shows Comparison among four categories of deep learning

Properties	RBMs	Sparse Coding	CNNs	Auto-Encoder
Generalization	✓	✓	✓	✓
Unsupervised learning	✓	✓	✗	✓
Feature learning	✓	✗	✓	✓
Biological understanding	✗	✓	✗	✗
Small training set	✓	✓	✓	✓

Table 2 shows advantages and dis advantages of four categories in deep learning. Normally, Generalization is used to show effective approach in diverse media such as images, texts, and audio; 'Unsupervised learning' mention about ability for learning a deep model without supervisory a notation. 'Feature learning' is mention learn feature based on a dataset without training. 'Biological understanding' shows whether this approach is basically a biological basis or theoretical basis. 'Small training set' mention to capability of a small number of example to learn a deep model.

6. BAG-OF-WORDS (BoW) METHOD
One of the supervised models is the BoW. The human has great skills to reduce the computational cost of a data-driven visual

search by attention mechanism. Bottom-up and top-down are two ways to direct attention [56]. In a bottoms-up method, for deployment of visual attention detecting salient regions in an image is utilized. Past knowledge is made available using prior knowledge availability for a specific target to guide the visual attention utilized by top-down way.

The advantage of Bow can be divided into i) mainly not affected by the position of the object being destructed in an image; position the object's destruction in an image, ii) Fixed length vector regardless of the amount of detections, iii) Successful in tabulating images based on the object they contain, vi) Still needs more examination for large changes in the scale and viewpoint. Shape of visual word position is not obvious and Localization objects are poor in images and a disadvantage to BoW. The bag-of-words method has been recommended in domain challenge text recovery for proposed in the text recovery domain problem for text document analysis [57]. BoW model is utilized by visual analogue of a word that is developed in the process of the vector quantization by clustering low-level visual aspects of points or the localized regions including the texture as well as the color [58]. BoW includes some steps that extract the features as mentioned below:

(i)   Automatic detection of the points/regions of interest;

(ii) Compute the descriptors that are local across the points/region;

(iii)   Quantify the descriptors into words that outline the visual vocabulary; and

(iv) Look for the image's occurrences in each particular word in the vocabulary to construct the feature of the Bag of Words.

7. DISCUSSION

The color-based approach depends on the light of the activity suspicion that pixels in pornographic images are primarily skin. Color is the most straightforward visual quality to model, and it is a characteristic beginning stage when working with huge data sets. Three-dimensional color spaces bring about computationally reasonable calculations and algorithms that can be simply visualized. Global descriptor is used by shape. Texture, local pattern, and color have not been detecting the internal details with Shape. Thus, the classification of objects is used by shape. It cannot be utilized to recognize two objects of the same shape with various texture or colors. Therefore, you cannot utilize shape feature for object recognition.

Shape-based recovery will likewise basically represent the query and database images in some feature space, and endeavor Map and Reduce operations for discovering separations from the query image and restore the cases with minimum distance.

For improving the detection performance, one typically uses shape and skin color together. Global feature is utilized for low-level application such as classification and object detection. Local feature is applied for higher level application for example object recognition. To improve the accuracy of recognition, the Integration for local and global features was used.

SIFT computes the gradient histogram only for patches (usually 16*16 divided into 16 cells) around specific interest points obtained by taking the DoG's (as an approximation to LoG's) in the scale space. Its a local descriptor. SIFT in its original version by David Lowe is a 128 dimensional vector. SIFT features are usually compared by computing the Euclidean distance between them. the pLSA model is extended by sampling those weights from a Dirichlet distribution. This extension allows the model to assign probabilities to data outside the training corpus and uses fewer parameters, which can reduce the overfitting problem. Its main goal is to model cooccurrence information under a probabilistic framework in order to discover the underlying semantic structure of the data.

Neural network is interpreted by these architectures. Level of network hardwiring is the Key difference between auto-encoder and CNN. Convolutional nets are hardwired. Convolution operation much have a connection view by sparsity in the number of connection in the neural network. Pooling (subsampling) operation in image space is likewise a hardwired set of neural connection in the neural domain. Usually, the CNN is used for image and speech task utilized.

Particularly the topology of the network is not specified by auto-enders. To rebuild the input, it is good knowledge to find great neural transformation. The decoder and encoder compos reconstruct the input. Latent factors or latent features are learned by a hidden layer.

Interpretation of the neural network is different with RMB. A two-part graph is interpreted by RMBs where knowledge is the learning from the distribution of hidden and input variables. CNN and Auto encoder have learned a function. Additionally, the model is generative by RBMs. The sample is created from a learned hidden representation. RBMS are trained by different algorithms. However, at the end of the day, after learning RBMs, you can use its network weights to interpret it as a feedforward network.

Sparse coding is taught to a general class of methods that automatically select a sparse arrangement of vectors from the expansive pool of conceivable bases to encode an input signal. Sparse coding can simply be connected to the BoW framework as a substitution for vector quantization. This approach claims that the utilization of sparse coding to develop abnormal state highlights, demonstrating that the resulting representation performance is better than the conventional representation. Utilization the sparse coding to develop abnormal state highlights demonstrates that the resulting representation performance is better than the conventional representation.

8. CONCLUSION

This paper presents a review of methods for flirting porn image. It divides them into six categories: color base which is color is the most straightforward visual quality to model, and it is a characteristic beginning stage when working with huge data sets. Shape based recovery will likewise basically represent the query and database images in some feature space, local and global, feature approach is utilized for low-level application such as classification and object detection. Bag of words is utilized to image classification, treating image features as words and deep which is utilized Computational model by deep learning to compose multiprocessing layers to learn representations of data with multiple levels of abstraction. The paper mainly reports the developments of the filtering methods. Each of them is included in some categories. The advantage and disadvantage are discussed in this paper.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] X. Wang, C. Hu, and S. Yao, "A breast detecting algorithm for adult image recognition," in Information Management, Innovation Management and Industrial Engineering, 2009 International Conference on, 2009, pp. 341-344.

[2] M. N. Patel and P. Tandel, "A Survey on Feature Extraction Techniques for Shape based Object Recognition," Image, vol. 137, 2016.

[3] C.-Y. Jeong, J.-S. Kim, and K.-S. Hong, "Appearance-based nude image detection," in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004, pp. 467-470.

[4] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," IEEE transactions on pattern analysis and machine intelligence, vol. 24, pp. 696-706, 2002.

[5] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," Image and vision computing, vol. 21, pp. 99-110, 2003.

[6] L. Duan, G. Cui, W. Gao, and H. Zhang, "Adult image detection method base-on skin color model and support vector machine," in Asian conference on computer vision, 2002, pp. 797-800.

[7] J. Kovac, P. Peer, and F. Solina, Human skin color clustering for face detection vol. 2: IEEE, 2003.

[8] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," International Journal of Computer Vision, vol. 46, pp. 81-96, 2002.

[9] Y. Huang and A. W. K. Kong, "Using a CNN ensemble for detecting pornographic and upskirt images," in Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on, 2016, pp. 1-7.

[10] N. Sae-Bae, X. Sun, H. T. Sencar, and N. D. Memon, "Towards automatic detection of child pornography," in Image Processing (ICIP), 2014 IEEE International Conference on, 2014, pp. 5332-5336.

[11] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," Pattern Recognition, vol. 46, pp. 188-198, 2013.

[12] G. Pass and R. Zabih, "Comparing images using joint histograms," Multimedia systems, vol. 7, pp. 234-240, 1999.

[13] D. Zhang and G. Lu, "Review of shape representation and description techniques," Pattern recognition, vol. 37, pp. 1-19, 2004.

[14] W. Hsu, S. Chua, and H. Pung, "An integrated color-spatial approach to content-based image retrieval," in Proceedings of the third ACM international conference on Multimedia, 1995, pp. 305-313.

[15] Q.-F. Zheng, W. Zeng, W.-Q. Wang, and W. Gao, "Shape-based adult image detection," International Journal of Image and Graphics, vol. 6, pp. 115-124, 2006.

[16] R. Lienhart and R. Hauke, "Filtering adult image content with topic models," in Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, 2009, pp. 1472-1475.

[17] J. Shotton, A. Blake, and R. Cipolla, "Contour-based learning for object detection," in Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, 2005, pp. 503-510.

[18] P. S. Nikkam and E. B. Reddy, "A Key Point Selection Shape Technique for Content based Image Retrieval System," International Journal of Computer Vision and Image Processing (IJCVIP), vol. 6, pp. 54-70, 2016.

[19] C. Tian, X. Zhang, W. Wei, and X. Gao, "Color pornographic image detection based on color-saliency preserved mixture deformable part model," Multimedia Tools and Applications, pp. 1-17, 2017.

[20] C. Tian, X. Zhang, W. Wei, and X. Gao, "Color pornographic image detection based on color-saliency preserved mixture deformable part model," Multimedia Tools and Applications, pp. 1-17, 2017.

[21] S. S. Chaeikar, M. Zamani, A. B. A. Manaf, and A. M. Zeki, "PSW statistical LSB image steganalysis," Multimedia Tools and Applications, pp. 1-31, 2016.

[22] C. Zhai, "Probabilistic Topic Models for Text Data Retrieval and Analysis," in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 1399-1401.

[23] X. Wu, Y. Tang, and W. Bu, "Offline text-independent writer identification based on scale invariant feature transform," IEEE Transactions on Information Forensics and Security, vol. 9, pp. 526-536, 2014.

[24] Y. Raoui, E. H. Bouyakhf, M. Devy, and F. Regragui, "Global and local image descriptors for content based image retrieval and object recognition," Applied Mathematical Sciences, vol. 5, pp. 2109-2136, 2011.

[25] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," Machine learning, vol. 42, pp. 177-196, 2001.

[26] J. Wang, P. Liu, M. F. She, A. Kouzani, and S. Nahavandi, "Supervised learning probabilistic latent semantic analysis for human motion analysis," Neurocomputing, vol. 100, pp. 134-143, 2013.

[27] D. Oneata, "Probabilistic Latent Semantic Analysis," in Proceedings of the Fifteenth conference on Uncertainty, 1999, pp. 1-7.

[28] T. Lindeberg, "Scale invariant feature transform," Scholarpedia, vol. 7, p. 10491, 2012.

[29] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 2528-2535.

[30] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 49, pp. 4516-4527, 2011.

[31] W. Burger and M. J. Burge, "Scale-Invariant Feature Transform (SIFT)," in Digital Image Processing, ed: Springer, 2016, pp. 609-664.

[32] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85-117, 2015.

[33] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 2012, pp. 17-36.

[34] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," arXiv preprint arXiv:1405.3531, 2014.

[35] 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.

[36] S. Yazdanpanah, S. S. Chaeikar, M. Zamani, and R. Kourdi, "Security features comparison of master key and IKM cryptographic key management for researchers and developers," in International Conference on Software Technology and Engineering, 3rd (ICSTE 2011), 2011.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

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

[39] M. Lin, Q. Chen, and S. Yan, "Network in network," arXiv preprint arXiv:1312.4400, 2013.

[40] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision, 2014, pp. 818-833.

[41] R. Alvear-Sandoval and A. R. Figueiras-Vidal, "Does diversity improve deep learning?," in Signal Processing Conference (EUSIPCO), 2015 23rd European, 2015, pp. 2496-2500.

[42] C.-Y. Liou, W.-C. Cheng, J.-W. Liou, and D.-R. Liou, "Autoencoder for words," Neurocomputing, vol. 139, pp. 84-96, 2014.

[43] C. Poultney, S. Chopra, and Y. L. Cun, "Efficient learning of sparse representations with an energy-based model," in Advances in neural information processing systems, 2007, pp. 1137-1144.

[44] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in Proceedings of the 25th international conference on Machine learning, 2008, pp. 1096-1103.

[45] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in Proceedings of the 28th international conference on machine learning (ICML-11), 2011, pp. 833-840.

[46] H. Lee, C. Ekanadham, and A. Y. Ng, "Sparse deep belief net model for visual area V2," in Advances in neural information processing systems, 2008, pp. 873-880.

[47] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 1794-1801.

[48] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in Advances in neural information processing systems, 2009, pp. 2223-2231.

[49] X. Zhou, K. Yu, T. Zhang, and T. S. Huang, "Image classification using super-vector coding of local image descriptors," in European conference on computer vision, 2010, pp. 141-154.

[50] S. Gao, I. W.-H. Tsang, L.-T. Chia, and P. Zhao, "Local features are not lonely–Laplacian sparse coding for image classification," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 3555-3561.

[51] S. S. Chaeikar, A. B. A. Manaf, and M. Zamani, "Comparative analysis of Master-key and Interpretative Key Management (IKM) frameworks," in Cryptography and security in computing, ed: InTech, 2012.

[52] G. E. Hinton and T. J. Sejnowski, "Learning and releaming in Boltzmann machines," Parallel Distrilmted Processing, vol. 1, 1986.

[53] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural computation, vol. 18, pp. 1527-1554, 2006.

[54] R. Salakhutdinov and G. Hinton, "Deep boltzmann machines," in Artificial Intelligence and Statistics, 2009, pp. 448-455.

[55] J. Ngiam, Z. Chen, P. W. Koh, and A. Y. Ng, "Learning deep energy models," in Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, pp. 1105-1112.

[56] C.-F. Tsai, "Bag-of-words representation in image annotation: A review," ISRN Artificial Intelligence, vol. 2012, 2012.

[57] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in Computer vision and pattern recognition, 2006 IEEE computer society conference on, 2006, pp. 2169-2178.

[58] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in Workshop on statistical learning in computer vision, ECCV, 2004, pp. 1-2.