International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Dense Neural Patterns and Innovative CNN Architectures for Efficient Object Detection and Image Classification

Bheesetty Srinivasa Rao

Submitted:15/05/2024 **Revised:** 27/06/2024 **Accepted:** 05/07/2024

Abstract: This paper presents a comprehensive study on image classification and object detection, showcasing significant advancements in these domains. For object detection, we proposed an innovative approach by integrating Dense Neural Patterns (DNPs) with traditional detection frameworks, derived from deep convolutional neural networks (CNNs). This method notably improved performance in the Regionlets detection framework, achieving superior results on the PASCAL VOC datasets. In the realm of image classification, we introduced several key innovations: the Latent CNN, designed to manage multi-label images by focusing on discriminative regions; Multiple Instance Learning Convolutional Neural Networks (MILCNN), which enhance deep learning capabilities with limited labeled data; and the Residual Networks of Residual Networks (RoR) architecture, aimed at improving optimization. Despite these advancements, we identified areas for further improvement, such as speeding up detection through CNN-based bounding box proposals, advancing unsupervised learning techniques, and using RNNs with LSTM units to generate more effective image regions for classification.

***Keywords**: CNN, RNN, SVM, Object detection.*

1. Introduction

Deep learning methods are revolutionizing the field of image classification and detection. The breakthrough in Image Net challenge [23] has demonstrated how powerful feature representations can be learned from data automatically, outdating traditional approaches based on hand-designed features. The most successful algorithm in a deep learning field has convolution neural networks (CNN), which combine three architectural ideas to ensure some degree of shift and distortion invariance: local receptive fields, shared weights, and spatial sub sampling. Recently, the success of CNNs has been attributed to their ability to learn rich high-level image representations as op- posed to hand-designed low-level features used in other image classification methods. Dense Neural Patterns (DNP) [24] have demonstrated that deep CNN features are substantially different from and complementary to those traditional features used in object detection. However, searching the parameter space of deep architectures is a difficult task because the training criterion is non-convex and involves many local minima. Many techniques, such as Relu [25], Dropout [26], Drop connect [27], pre- training [28] and data augmentation [29], have been proposed to enhance the performance of deep architectures. Despite local minima problems, deep convolution neural networks recently achieved remarkable successes in many visual recognition tasks, such as image classification and object detection, fine grained recognition, and visual instance retrieval [30].

The quality of visual features is crucial for image classification and detection. In the last decade, considerable progress has been made on the hand-designed features, such as SIFT [7], HOG [8] and LBP [9], and manually designed sophisticated coding schemes, such as BoF [10], SPM[11]. Recently, meaningful high-level features learned from CNNs achieve remarkable successes and the focus of visual object recognition research is shifting from feature engineering to deep network design and optimization.

Research Scholar, Department of Computer Science, Dravidian University, Kuppam.
Working at GITAM Deemed to be university, Department of Computer science, GITAM School of science.
srinivas66022838@gmail.com / bheeset@gitam.edu

Contributions

The primary objective of this research is to develop algorithms for learning powerful features by optimizing a deep convolutional neural network. In this dissertation, we make the following contributions:

Deep Neural Patterns

We presented a novel framework to incorporate a discriminatively trained deep convolutional neural network into generic object detection. It is a fast effective way to enhance existing conventional detection approaches with the power of a deep CNN. Instantiated with Region lets detection framework, we demonstrated the effectiveness of the proposed approach on public benchmarks. We achieved comparable performance to state-of-the-art with 74 times faster speed on PASCAL VOC datasets. We also showed that the DNPs are complementary to traditional features used in object detection. Their combination significantly boosts the performance of each individual feature.

Latent Model Classification

We developed Latent CNN to handle images with multiple labels during the training procedure. Latent CNN is designed to automatically select the most discriminate region to reduce the effect of irrelevant regions. We also proposed a new combination scheme for multiple CNNs called Latent Model Ensemble in order to reduce the local minima effect of deep CNNs.

Multiple Instance Learning Convolution Neural Networks

We provided a weakly supervised framework for image recognition by combining multiple instance learning loss and deep residual networks. We presented a mathematical formulation for how to incorporate the concept of multiple instance learning to a deep learning architecture, and we showed state-of-the-art performance on both low- resolution CIFAR datasets and high-resolution ILSVRC2015 classification dataset.

2. Literature Review

The architecture of typical deep CNNs is a stack of convolution, non-linear, pooling and fully-connected layers, followed by a loss function layer. It is designed to take advantage of local connections, shared weights, pooling, and the use of many layers to learn high-level representations of natural images, and it has demonstrated significant improvement over various benchmark object recognition datasets [12].

In the previous quite a while, more profound and more profound CNNs have been built on the grounds that the more convolutional layers and better improvement ability are in CNNs. From 5-conv+3-fc AlexNet to the 16-conv+3-fc VGG networks and 21-conv+1-fc Google Net(ILSVRC2015winner), both the exactness and profundity of CNNs were expanding. Nonetheless, profound CNNs confront a urgent issue, vanishing slopes. Prior works received instatement strategies and layer-wise preparing to decrease this issue. In addition, ReLU initiation work and its variations ELU, PReLU, PELU additionally can avert vanishing angles. Luckily, this issue can be to a great extent tended to by group standardization (BN) and painstakingly standardized weights introduction as per late research. BN institutionalized every short cluster with mean and fluctuation of shrouded layers, while MSR introduced the weights by more sensible change. In another angle, a corruption issue has been uncoveredthat is, not all frameworks are correspondingly simple to improve. With a specific end goal to determine this issue, a few techniques were proposed. Expressway Networks comprises of an instrument al-lowing 2D-CNNs to associate with a basic memory component. Indeed, even with several layers, expressway systems can be prepared specifically through basic angle drop. ResNets[13-15] improved Highway Networks utilizing a simples kip association component to proliferate data to further layers of systems. ResNets are turned out to be simpler to utilize and more successful than thruway Networks. As of late, FractalNet produced a to a great degree profound system whose basic design was filled in as an exact truncated fractal by rehashing utilization of a solitary extension manage, and this strategy demonstrated that leftover learning was not required for ultra-profound systems. Nonetheless, keeping in mind the end goal to get the aggressive execution of ResNets, FractalNet requires numerous a greater number of parameters than ResNets. Presently, more Residual system variations and designs have been proposed, and they shape a leftover systems family together.

3. Methodology

Visual Analysis

Is the expansion in location execution by adding thick neural examples inferable from the abnormal state signals encoded by DNPs? To answer this inquiry, we devise representation strategies for the most imperative highlights utilized by the indicator. The learning procedure for boosting chooses discriminative powerless classifiers. The significance of an element measurement generally compares to how much of the time it is chosen amid preparing. We check the event of each measurement in the last feeble classifier set and discover the DNP include measurement most often chose by boosting. To envision these element measurements, we recover picture crops from the dataset which give the most astounding reactions to the comparing neurons in the profound CNN.

Figure 3.8 demonstrates the perception. The perfect case is that the most continuous neural examples chose in a man locator give high reactions to parts having a place with a man. This shows the neural examples encode abnormal state data. The left section of Figure 3.2 portrays the protest class we need to distinguish. Right segments indicate visual patches which give high reactions to the most every now and again chose neural example measurement for the class. This examination demonstrates that the chose neural examples encode part-level or question level visual highlights exceedingly connected with the protest classification. For a canine finder, neural examples identified with a puppy confront are as often as possible chose. We likewise played out a comparable examination with the HOG include. In comparison, the frequently selected HOG dimension carries a lot less categorical information because gradients are low-level visual features.

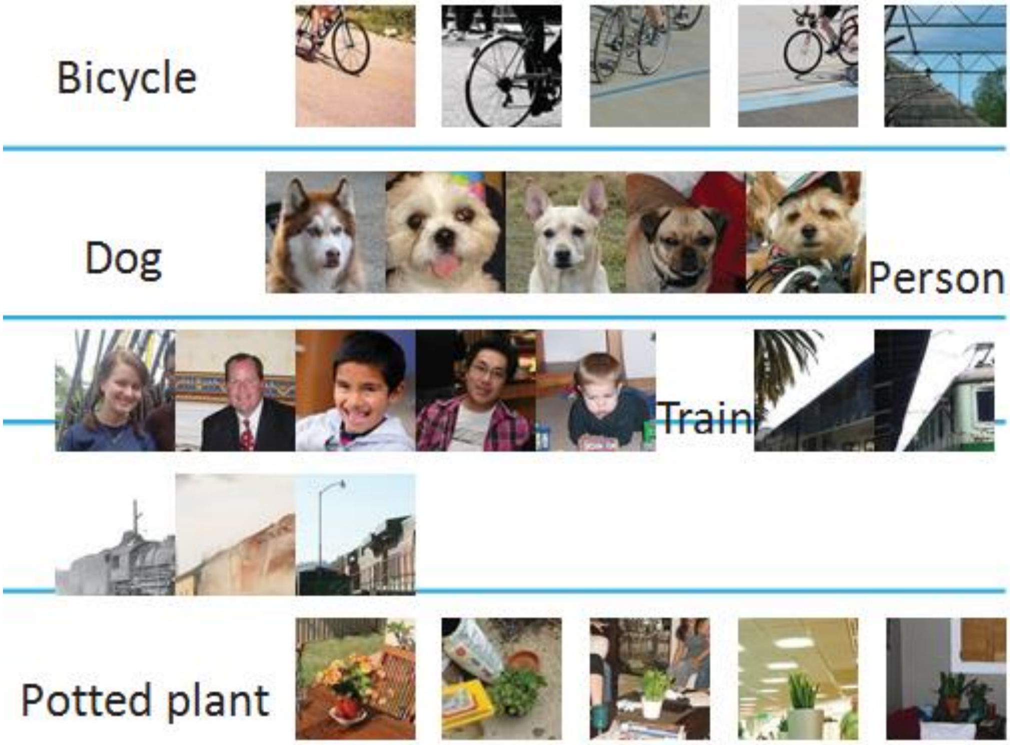

Figure 3.1: Visualization of the abnormal state data encoded by neural examples from the fifth convolutional layer.

Deep CNNs

Alex Net is a pile of convolutional layers, which are alternatively trailed by differentiate standardization layers and max-pooling layers, and privately associated layers or completely associated layers. Exchange learning is received when the Alex Net is utilized for other littler dataset. For instance, the PASCALVOC 2007 order dataset has just 5,011 preparing pictures, which makes it practically difficult to take in a fulfilled profound CNN model.

Our CNN structure prepared on Image Net is like, which has five convolutional layers, two completely associated layers, and one yield layer. The maximum pooling layers are added to conv1, conv2, and conv5 layer, separately. The distinction is that normalizes every one of the pictures into 256x256 squares and takes 224x224 fixes as the in-put for the CNN structure. Our CNN structure, in any case, does not have to standardize the first pictures and it takes 128x128 patches. Contrasted with Alex Net picture interpretation strategy, our technique may not be mark saving changes, looking at that as a considerably littler area most likely does not fulfill the half cover run the show. Be that as it may, it isn't an issue when it is connected to the dormant CNN structure, which would choose the most discriminative area as the contribution for CNN learning method. Exchange learning is utilized for preparing the CNN structure for the PASCAL VOC dataset. Contrasted with the CNN show prepared from Image Net, it evacuates the last1000-hub yield layer and includes two all the more completely associated layers. Amid the preparation for the PASCAL VOC 2007 order dataset, all the exchanged parameters are settled first and final the parameters in the last two completely associated layers are refreshed.

At last, an adjusting of the entire CNN structure is connected.

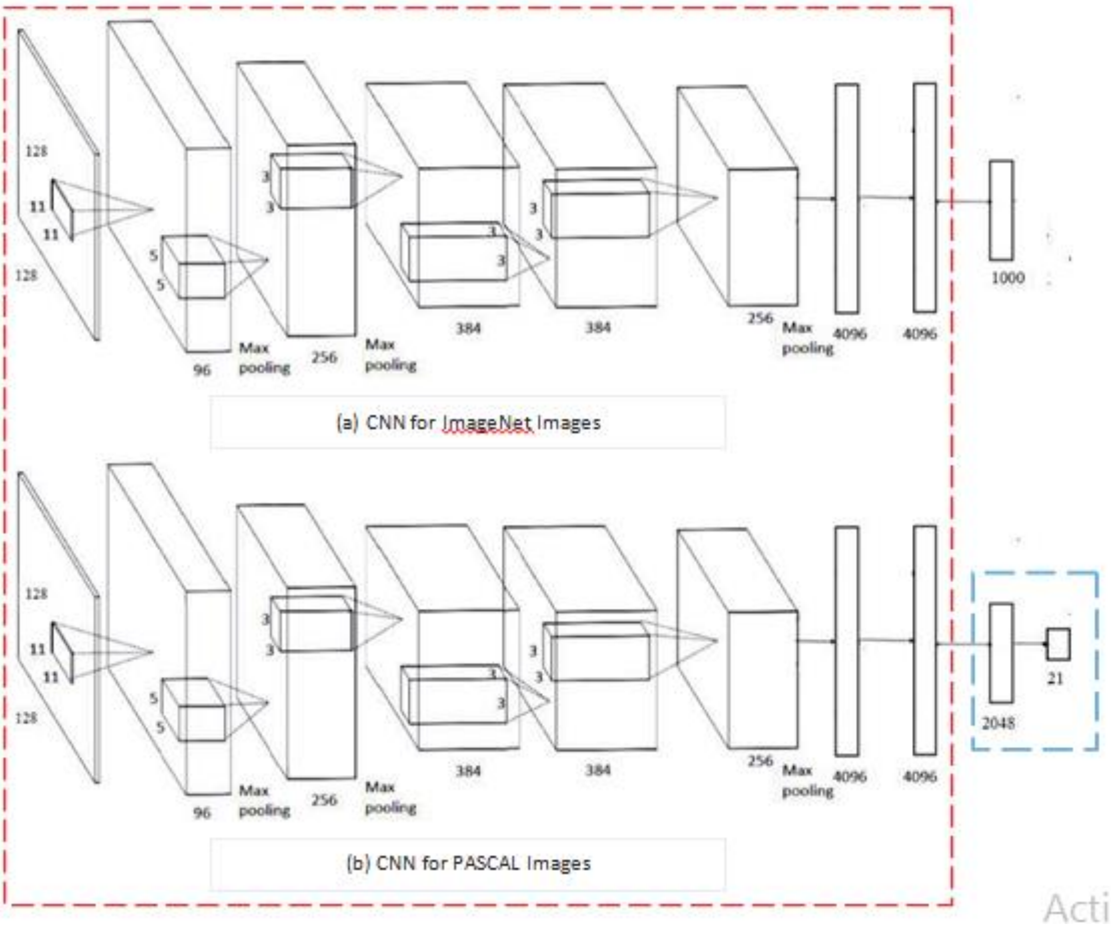

Figure 3.2: Transfer learning. (an) is the CNN gained from image net pictures while (b) is the CNN utilized for PASCAL pictures. The red rectangular signifies the parameters exchanged from CNN (a) to CNN (b). When preparing the CNN in (b), just the parameters in the blue rectangular boxes are refreshed.

4. Results And Study

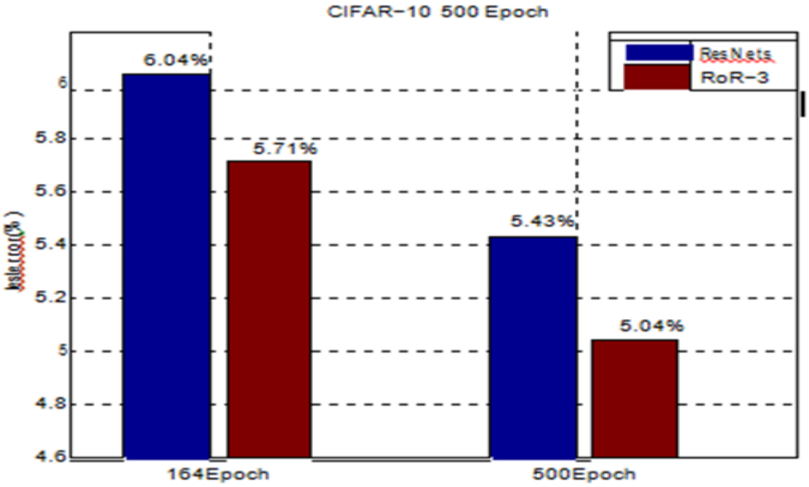

Figure 4.1: CIFAR-10 - Comparison of ResNets and RoR-3 with various age numbers

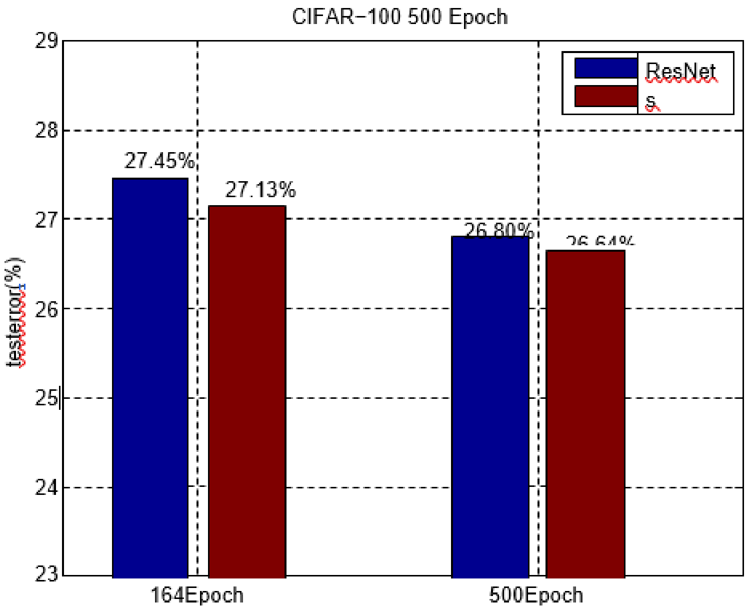

Figure 4.2: CIFAR-100 - Evaluation of ResNets and RoR-3 with various age numbers.

Fig. 4.1, Fig. 4.2 will demonstrate the preparation five zero zero (500) ages could become critical advancement. So could pick five zero zero (500) by means of the most extreme age number.

Table 4.1: Test Error on ResNets RoR and CIFAR-10

(CIFAR-10) 500 Epoch	ResNets	(ResNets) + SD	RoR-3	(RoR-3) + SD
110-layer	5.43	5.63	5.08	5.04
164-layer	5.07	5.06	4.86	4.90

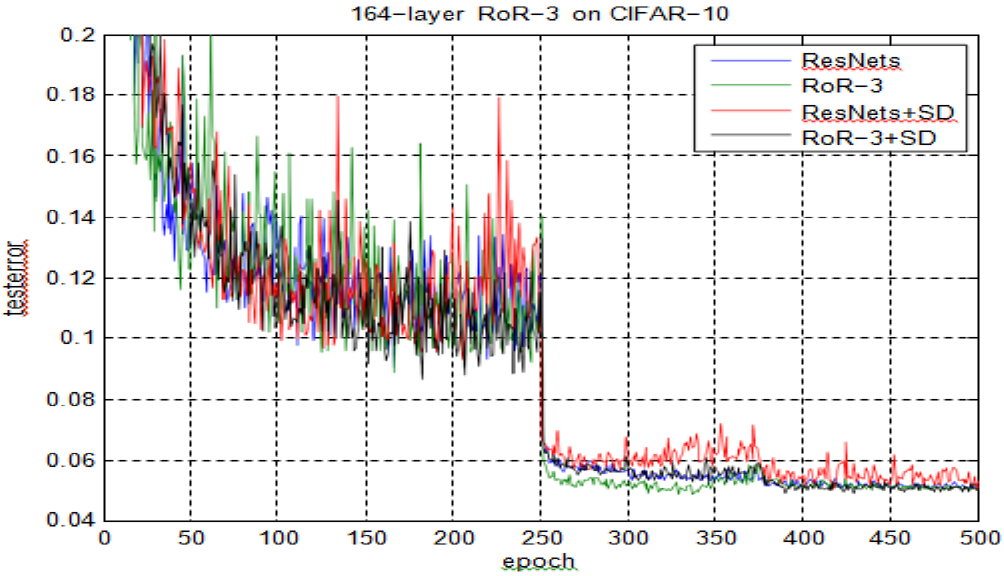

Figure 4.3: CIFAR-10 - Test mistakes by ResNets, RoR-3, (ResNets)+SD and (RoR-3)+SD

In Table 4.1 and Fig. 4.3, 110-layer ResNets without SD brings about an aggressive 5.43% mistake on the test set. 110-layer RoR-3 without SD brings about a 5.08% blunder on the test set, and it outflanks 110-layer ResNets without SD by 6.4% on CIFAR-10 with the comparative number of parameters. 164-layer RoR-3 without SD brings about a 4.86% mistake on the test set, and it outflanks 164-layer ResNets without SD by 4.1%.

Conclusion

This paper encapsulates our exploration into image classification and object detection. For object detection, we tackled the challenge of integrating deep convolutional neural networks with traditional detection frameworks, resulting in the development of Dense Neural Patterns (DNPs). These features, derived from deep CNNs, were effectively applied within the Regionlets framework, significantly enhancing performance on the PASCAL VOC datasets. In the domain of image classification, we achieved notable progress by developing the Latent CNN, which adeptly manages multi-label images by selecting the most discriminative regions. Additionally, we introduced Multiple Instance Learning Convolutional Neural Networks (MILCNN) to maximize deep learning's potential despite limited labeled data, and the Residual Networks of Residual Networks (RoR) architecture to improve the optimization capabilities of residual networks.

References

[1] Subhransu Maji, Alexander C. Berg, and Jitendra Malik. Classification using intersection kernel support vector machines isefficient.

[2] Chaitanya Desai, Deva Ramanan, and Charless C. Fowlkes. Discriminative models for multi-class object l ayout. International Journal of Computer Vision, 95(1):1–12,2011.

[3] Pablo And rés Arbeláez, Jordi Pont-Tuset, Jonathan T. Barron, Ferran Marqués, and Jitendra Malik. Multiscale combinatorial grouping. In CVPR, 2014.

[4] Ming ming Cheng, Ziming Zhang, Wen yan Lin, and Philip Torr. Binarized normed gradients for objectness estimation at 300fps. In in IEEECVPR, 2014.

[5] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. Imagenet large scale visual recognition competition 2012(ilsvrc2012). 2012.

[6] Will Y. Zou, Xiaoyu Wang, Miao Sun, and Yuanqing Lin. Generic object detection with dense neural patterns and regionlets. In BMVC,2014.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In Neural Information Processing Systems, NIPS2012.

[8] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Improving neural networks by preventing co-adaptation

of feature detectors. In CoRR,2012.

[9] W. Li, M. Zeiler, S. Zhang, Y. LeCun, and R. Fergus. Regularization of neural networks using dropconnect. In ICML,2013.

[10] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pre-training help deep learning? In JMLR, volume 11(Feb), page 625660,2010.

[11] J. Schmidhuber. Multi-column deep neural networks for image classification. In CVPR,2012.

[12] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In CVPR workshop,2014.

[13] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge, 2010.

[14] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index. html.

[15] Fahad Shahbaz Khan, Rao Muhammad Anwer, Joost van de Weijer, Andrew D. Bagdanov, Marı́a Vanrell, and Antonio M. Ló́pez. Color attributes for object detection. In CVPR,2012.

[16] Timo Ahonen, Student Member, Abdenour Hadid, Matti Pietikinen, and Se- nior Member. Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28:2037–2041,2006.

[17] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan. An hog-lbp human detector with partial occlusion handling. In ICCV,2009.

[18] Oncel Tuzel, Fatih Porikli, and Peter Meer. Pedestrian detection via classi- fication on riemannian manifolds. IEEE Trans. Pattern Anal. Mach. Intell., 30(10):1713–1727,2008.

[19] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce. Learning mid-level features for recognition. In CVPR,2010.

[20] M. A. Wong J. A. Hartigan. Algorithm as 136: A k-means clustering algorithm. Applied Statistics, 28(1):100–108, 1979.