



MATERIAL CLASSIFICATION SYSTEM: LITERATURE SURVEY

Shama Holla
Department of Computer
Science,
BMS College of
Engineering
Bangalore, Karnataka,
India.

Shivani Bonageri
Department of Computer
Science,
BMS College of
Engineering
Bangalore, Karnataka,
India.

Shravya Shetty
Department of Computer
Science,
BMS College of
Engineering
Bangalore, Karnataka,
India.

K Panimozhi
Department of Computer
Science,
BMS College of
Engineering
Bangalore, Karnataka,
India.

Abstract— One of the chief challenges faced today is determining the material category of a surface from an image. From the results obtained from several studies on object detection and image classification using Convolutional Neural Networks (CNNs), it is possible to study the material classification of everyday objects. The following intensive literature survey enables us to understand the various techniques developed to overcome the challenges.

Keywords— CNN (convolutional neural networks), Deep Learning, Material Recognition, Image classification

I. INTRODUCTION

Material classification is a topic has received attention in recent times and a very few of the computer vision systems have been constructed to recognize materials in the world so far but the accuracy level isn't high as of now.

Image classification and object detection has been active areas of research in the last few years. Initially, approaches, such as Bagof-Visual-Words (BoVW), were employed that yielded fairly good results for these two tasks. However, the emergence of Convolution Neural Networks (CNNs) for solving these vision based problems has changed the scenario altogether by comprehensively outperforming the handcrafted approaches.

Throughout this survey, we have been able to understand the recent advances in the field and incorporate the progress made into our implementation to recognise materials from images.

II. LITERATURE SURVEY

It is often very difficult to numerically measure the relations between structure and property of objects made of composite material. The classification of materials that are not crystalline is very difficult as a result of their short-range order and

causes difficulty to outline measurements related to their structure. During the course of this experiment, we make use of deep learning algorithms to precisely group non-crystalline materials and identify measurements regarding their physical design. Particularly, we explain that message passing neural networks and CNN will categorize liquids with two dimensions and glasses that are liquid-cooled from molecular dynamics simulations. Without any beforehand suppositions regarding relationship between the native particles, even once the materials taken square measure ready at a similar existing energy. Moreover, we try to interpret that message passing neural networks outperform CNNs here in this context in every measurement and illustrability. We tend to draw out a neat understanding of how they value non-crystalline materials by employing a self-cautious procedure. Misapplication of the technique, we tend to obtain three important measurements that clearly describe formation of the glass. The strategies used here offer a mechanism to spot vital arrangement options in materials that would be lost by normal methods and gives us a novel view into the information obtained through neural networks. [1]

Different objects tend to show distinguishable profiles in terms of spectral lines. This enables approximation of the fabric constitution when its hyperspectral image is first obtained, and lateral database having spectral profiles is employed to filter it. This existent technique is quite useless, there as on being it is only a collection having projections of the non heritable metrics contributing to the task. A camera that is uniquely programmed having the capability of manufacturing pictures with associate degree spectral filter is employed. Filtering of spectral lines of the current image with the database of profiles is optically implemented using this camera. Speed is one of the advantages obtained. Having coaching information, there is a tendency to apply a spread of classical and trendy procedures as well as cnns to spot the



database of spectral profiles that enable classification of these materials. We have a tendency to verify the strategy in simulations on customary information sets also as real data employing a research laboratory image of the camera [2].

An object's degree of fragility and its acceptable use can be understood by identifying the object's composition. Sense modality sensing is one of the techniques used in many of the previous works related to this. Throughout the working of this paper, we've got an inclination to explore how the material composition of objects are estimated by robots using chemical analysis. It can be demonstrated that, for material recognition, several edges are provided by spectrometers. Moreover, they would prefer contact that is direct with an object. Commercially available spectrometers are employed to obtain different spectral measurements and a dataset of the results collected is formed. It can also be shown that the neural networks can precisely analyse the results obtained. The challenge to generalize material identification to new items is bigger and an accuracy of 79.1% is achieved. Also, we've got an inclination to demonstrate but a PR2 mechanism can make use of spectrometers for the approximation of the composition of everyday household items. As a result of this research, we have a tendency to discover that chemical analysis can be a fruitful approach for classification of materials throughout robotic manipulation.[3]

One of the most talked about topics relating to remote sensing is hyperspectral image (HSI) classification. The data produced by this technique is advanced in nature and helps in rightful categorization which may seem tough with old machine learning methods. The relation between the materials and the corresponding obtained data is non-linear and is dealt using this type of imaging sometimes. The problems associated with non-linearity can be effectively addressed using many deep learning methods that extract features accurately. HSIs are classified using deep learning, for which these applications have been a motivation. Classification of HS images faces a number of challenges. Challenges that cannot be successfully overcome using the existing techniques are briefed. This in turn introduces the benefits of deep learning to manage these difficulties. A framework is created for dividing the respective works into spectral-spatial, spatial- feature and spectral-feature networks. These are then used to correctly test the new accomplishments in HIS classification based on deep learning. Lastly, several classification ways which are deep learning-based are conducted one by one on real HSIs in experiments.[4]

For portable computer vision to enable its usage "in the wild", it has become necessary for the recognition of materials for real-world out of doors upper appearance. There has been a transition from reflectivity modelling pattern to representations based on image for procedure surface modeling which is the basis of material recognition. Amorein-

between approach is needed which makes use of every created radiometric cues and versatile capture of images. A framework can be developed for differential angular imaging. In the capture of images, even small differences give a redoubled look illustration and very important recognition improvement. There is a need to create a material related data which is large in scale and (GTOS) data. This data can be used towards autonomous agents. Over thirty thousand footage relating to around forty classes of out of doors tract beneath different physical conditions is contained in the data. The conditions include varied lighting and weather conditions. This large scale dataset can be leveraged using an approach known as a Differential Angular Imaging Network (DAIN).The different properties of the materials are to be extracted along with angular and abstraction gradients. The results prove to be better than single scan or Multi view quantity footage. Hence the results obtained show the usefulness of DAIN as a way for recognition of materials versatile in nature.[5]

The interactions with objects can be guided by referring to the properties of materials. Absorbing materials that are sleek in nature, as an example, would require fragility in handling, and care. One of the properties that could suit material used in bedding is softness. These properties may seem to not appear visually (for example softness). Their presence visually still has to be inferred. Such properties which we try to grasp about are known as material visual attributes. Important and useful knowledge for general scene learning can sometimes be obtained by extracting these attributes in pictures. These attributes also find measurements of native properties with no special extent or mounted type along with distinction of already known item and scene attributes. Native image packets are used to acknowledge the attributes given several footage combined with known attributes. However it is very troublesome to get this correctly and consistently. A new approach is introduced which examines human observation by asking very simply yes or no queries. It gives us adequate direction to form a gaggle of attributes and classifiers associated with them. Attributes of materials that are visual can be acknowledged without having to do manual annotation of a set of attributes. The results of the experiment prove that, visual material attributes, irrespective of whether they are named or unnamed or discovered automatically, are a good in between illustration.[6]

Material recognition is one of the significant open issues in computer vision. A method has been proposed to solve this. A time-of-flight camera has been employed to analyse the properties and attributes of different objects made of different materials. A depth sensor provides certain data and it is obtained to start with and varied options of interest are taken by using different transforms like Hilbert, Fourier, chirp-z, karhunen-loeve and distinct cos. Training and testing data set are built using the made choices. This is in turn useful to feed the classifier (J48) which successfully achieves the recognition



of material step. The accuracy and correctness of this procedure has been rightfully tested and verified. The accuracies of the materials obtained are often sensible. Moreover, hardness and responsibility of the calculated choices can be increased when using a mix of many transforms in an experiment. But rates of prediction can be heavily impacted by shutter rates.[7]

Inferring the attributes and properties of materials and objects is one of the most important tasks of visual data. An object's causative history is one of such properties. It could relate to what the article has undergone in its past like, "this foil has been crushed", etc. A footage of objects made of completely differing materials is taken ('aluminum foil', 'chicken wire', 'gold foil', 'wax', 'cardboard', 'putty') and one of the various shape altering methods ('bent', 'folded', 'twisted', 'crumpled') are re applied. They obtain thirty images of each object/shape alteration combination (720 images) by various application of view point and shape transformation. Different groups of participants are formed and asked to 1) identify materials and their respective transformations 2) rate the footage according to the extent they belonged to each shape altering class 3) categorize the images into different shape alteration class. It can be inferred from the results that the participants are easily able to identify transformations by looking at the shapes of the object, with the accuracy changed by material of the object. This information from causative history proves that we can differentiate between the original and transformed characteristics of the article.[8]

Recognition of the materials in distinct pictures requires a method which uses neural networks and also can be used to train models which is used to get the properties of that particular image. In this paper, they have used "Convolutional Neural Network (CNN) techniques and Support Vector Machine(SVM)" classifiers to group the different type of substances and then to study the outcome obtained. Numerous databases of materials is used to gauge different kind of CNN architectures so that they will be able to get the appropriate method to take out feature of a material. The outcome come obtained from the four material dataset and 9 CNNs says that the implementation of a CNN bringing into play linear SVM increases mean average. Also, the analysis of the 9 completely different CNN architectures, which uses transfer learning rather than extracting the knowledge from the previous convolutional layer, enhances the accuracy of the developed system. From this paper we can highlight the fact that if we use the database that contains very large amount of images, then we can enhance both performance and accuracy of the system that we develop.[9]

Recognizing materials is a difficult task because of the great variation in look within different classes and also among different classes. Existing strategies uses large receptive fields

as an input to treat an implicit fusion of materials and context. In this paper, they show that recognizing materials strictly from their native look and then desegregating it places yields much better results. Objects are combined and the estimates obtained from the independent CNNs are then given to the present network. This particular method helps us get rid of the need of creating enormous amount of training data. Also, they perform an in depth analysis of the results of "context granularity", "spatial resolution", and also the "network level" where the context is introduced. By using this method we will be able to attain the accuracy, even though we use fewer training data when compared to the previous methodologies.[10]

"Data-augmentation" is very important to train the "neural networks" for "image classification". Recent analysis in image classification has incontestable improved performance by considering larger networks and higher resolution pictures. Data augmentation is generally adopted at training time to improve model generalization and scale back over fitting. Typical transformations include: random-size crop, horizontal flip and color interference. The accuracy is additionally improved by combining multiple data augmentations at test time, though this implies that several forward passes are needed to classify one image. Another performance-boosting strategy is to classify a picture by feeding it at multiple resolutions, once more averaging the predictions. Applying a Convolutional Neural Network (CNN) classifier to a picture typically needs to pre-process the image. One of the key steps involves choosing a rectangular region within the input image, which is termed as Region of Classification (RoC). The RoC is then extracted and resized to a square crop of a size compatible with the CNN. Employing p-pooling technique rather than average pooling to check resolutions considerably higher than the training resolution improves the performance of the network. Over the years, training and testing pre-processing procedures have evolved to enhance the performance of CNNs, however thus far they have been optimized on an individual basis. This separate optimization has led to a big distribution shift between training and testing regimes with a undesirable impact on the test-time performance of models. This downside is resolved by collectively optimizing the choice of resolutions and scales at training and test time, while keeping identical roc sampling. By adjusting the crop resolution and via an easy and light-weight parameter adaptation, we can improve the accuracy of normal classifiers considerably.[11]

Material classification in real-world environments may be a difficult downside because of the massive effect of lighting conditions and viewing on appearance of a material. Hence we need training dataset and also "intra-class variance" of materials to train classifier. We can determine what part "real-world training images" plays in material classification domain by checking if "synthesized data" will be able to deduce better



than “real–world data”. When experimented on three popular material databases, the results obtained tell us that the best “CNN architecture”, which is already trained, attains the mean average accuracy of 91.03%, when the materials are classified in “cross-dataset” situations. In this paper, they have showed that it is possible to segment the images, which are synthetic, perfectly without manually segmenting these synthetic images instead it is possible to segment them automatically. In this paper, they have used “cross- dataset analysis”, which helps to indicate how the synthetic and real world training datasets work in complicated situations, where real world test datasets are used. In this paper they have trained n linear SVM classifiers and considered only those groups which are common to real-world and artificial information for the analysis purposes. It is shown that, we will be able to get the better results for material classification if synthetic training datasets are used instead of real-world training data.[12]

Describing the features of the dataset always remains as the first method in creating a machine learning model. Feature selection methods are used to construct good prototype and also to attain model interpretability. Now dimensionality reduction techniques are in demand for manufacturing compact feature representations. These techniques accelerates the model construction as we will be using smaller dataset, which is obtained by applying the reduction techniques on a feature set so that we will get the effective representation of a particular feature. But these techniques transform the initial description of the options, and hence provide restricted interpretability. Apart from this approach, we can also use features election method for higher models. The procedure of choosing a smallest of initial variables in such a way that any model designed using the data which contains only those features that belongs to the above set, thus increasing the performance of the model is called feature selection. Feature choice helps to improve the performance of the model by eliminating unwanted options. In addition to this, it has another advantage of keeping the original feature representation, and hence providing higher interpretability. By combining crystallographic and geometric micro structural descriptors, it is possible to decide stress hotspots which are formed under uniaxial tensile deformation. Choosing of “feature selection method” has to be done very carefully because the chosen method should be capable of finding the dependency even if the attributes are correlated .[13]

While recognizing materials, sometimes material attributes offers a biased intermediate representation, especially when we identify from the local appearance of the material. All the previous methods used to recognize the features of the material separately, just before identifying the category it belongs to. But the search conducted on material perception and computer vision analysis on “object and place recognition” says that the features can be created as a by-

product during the class identification method. A carefully-selected set of material categories is used to deal with the problem of local material recognition and hierarchy is created depending on the observation of each material family. A many step procedure is used to obtain both segmentation and presence data efficiently for a group of pictures. In the very first step, each image is shown to multiple annotators and it is filtered. By visualizing the properties of perceptual material attributes, it is possible to examine how different techniques separate the materials. It can be shown that the semantically meaningful visual material trait fits well with the discovered attributes and it enables us to identify the previously hidden material categories.[14]

Recognizing the real-world materials which have properties like “surface texture, geometry and lighting conditions” is very difficult. Here they have used a large-scale, open dataset of materials, “Materials in Context Database (MINC)” which is combined with deep learning methods to correctly identify the material and group those images of the material. Training the system to identify at every pixel in an image of a material is the main aim. Process of training is divided into many phases and analysis of the performance is done in every phase. Firstly, CNN is trained which predicts for a given input patch. It is then used to predict materials on a dense grid across a picture. Using MINC database, it is possible to evaluate deep learning algorithms for classifying the material and thus obtain outcome that beats the previous procedures used for recognizing the materials.[15]

Classifying images is a important aspect of computer vision algorithms. A single gray scale image creates the hardest classification task. But colour images, hyper-spectral images and polarizing filters gives us massive details. In image segmentation major part of the work uses only a single image to do analysis. Some of the methods uses modified version of the previous one by collecting HDR or multispectral images. By using several pictures of a particular thing under different conditions of light we will be able to attain many samples of the surfaces of an object – “Bidirectional Reflectance Distribution Function (BRDF)”. It can be shown that this extra details will increase the category of material types which is classified by segmentation, and by using BRDF images we can further increase the precision without needing an explicit approximation of the material BRDF.[16]

In this work they have explored how the “convolutional network depth” has effect on its accuracy in the “image-recognition” setting in large-scale.

Large-scale image and video identification has become achievable due to Convolutional networks because of large repositories consisting of public images and computing systems like GPUs that has high-performance.



With ConvNets becoming very significant, several efforts have been put towards improving original architecture for obtaining better accuracy. In this paper, they have addressed an important characteristic of the ConvNet architecture design - its depth. In order to achieve this, rest of the parameters in architecture are fixed, and the depth is feasibly increased steadily in the network by addition of convolutional layers. In this paper they have come up with better ConvNet architectures that obtains higher accuracy with respect to "ILSVRC classification and localization tasks". The ConvNet architectures can also be used on other datasets that are devoted to image recognition.[17]

In this paper[18] a scalable approach called Concept Learner has been proposed to discover various visual concepts given from image collections that are weakly labeled. This paper shows that several detectors of visual concepts. They are learned without having human in loop and can be applied for the image-level identification of concepts and also for accurate identification of the concepts at "image region-level".

To discover various visual concepts given from image collections that are weakly labeled and for supervising the application of the concept classifiers that are weakly-learned on datasets; the max-margin algorithm is used. In the domain-selected supervision, evaluation of the concepts that are learned on benchmarks for object detection and scene recognition are done.

Concept Learner, which is a learning approach (hard instance) of the max margin to discover various visual concepts given from image collections that are weakly labeled is developed.

This paper[19] employs a technique that uses CNNs to recognize and extract spatial relations that is present from images. This paper has addressed two challenges:

Providing an understanding into how the network learns of the spatial relations Identifying the parts of an image that are used to predict the relations.

VGGNet, a network which is pre-trained, is employed to obtain features from image and to train "Multi-layer Perceptron (MLP)" on a collection of synthetic images. The sun09 dataset has been used for extraction of spatial relations. MLP is analysed in this paper to depict the relation between spatial relation prediction and consistent groups of nodes activation. How the network's capacity to identify the relations is affected due to the loss of these consistent groups.

Shading has been used to deal with the issue in the recovery of shape, illumination and albedo from only one grayscale image of a particular object. The construction of statistical models of shape and albedo and an optimization problem which searches for the most probable explanation of a single image is defined. The model developed, "SAIFS (shape, albedo, and illumination from shading)", gives plausible results on random grayscale images that are taken from real world.

The technique that has been used is focused on objects, that are to be extended, to scene Segmentation techniques can be used to generate candidate objects. Shading is the algorithm's primary prompt to infer depth, if there is a scene that is shading-free, the model reduces the scene into a "shape-from-contour" algorithm. The model developed outperforms all the algorithms published previously on the MIT Intrinsic Images dataset.[20]

In this paper [21] material recognition using a large set of local features to study real-world objects is analysed. This paper explores the various visual features and learning techniques for the same. The Kernel Descriptor frameworks been used and set of descriptors has been extended to include attributes that are material motivated using variances of magnitude and gradient orientation. the accuracy on the Flickr dataset has been improved from about 45% to 54%. ImageNet and macro photos has been used to introduce two new datasets which helps not only in the evaluation of the set of features but also for showing connections between object and material recognition.

The work in this paper helps make advances towards robust material recognition, its applications in real-world conditions. they have demonstrated that the same set classification techniques and local features work well for both of the recognition tasks, and also there are interactions when considering the two tasks together that is interesting, which is the foundation to explore multi-task learning, forming a base for future work.

Since the information describing the materials that make up scene constituents provides a better understanding of images it is optimal to acquire the material information at pixel level(every pixel),irrespective of the involved objects. In this paper, they have introduced visual material traits are employed to achieve this.

Material traits encode the material properties appearances.

Using learning of convolution kernels in an unsupervised setting helps to identify complex material traits that appears at every pixel and avoiding the influence from object-specific information is a part of this framework. Therefore, accurate recognition of material traits irrespective of the object manifesting them is possible. The outcomes in this paper show that material trait scan be accurately identified. it also includes a demonstration of the usage of material traits in image segmentation and material recognition.[22]

In this paper[23], they have evaluated to check if the features obtained from the "activation of a deep CNN" which is trained on fixed, large set of tasks dedicated to object recognition can be repurposed to new generic asks or not .investigation and visualization of the "semantic clustering of deep convolutional features" with respect to a range of such tasks is a part of this paper. Deep convolutional model is first trained in a



completely supervised setting employing method called Krizhevsky . Various features from this network is extracted and the efficiency of these features on generic vision tasks is evaluated. This paper also releases the “DeCAF”, which is the open-source implementation of the features, and also several associated parameters of network to enable the vision researchers in conducting experimentation across a wide scope of paradigms of visual concept learning.

This paper [24] proposes a simple detection algorithm that improves “mean average precision (mAP)” by more than 30% compared to the best result obtained in the past.

A mAP of 53.3% is obtained with this algorithm. The approach combines two aspects:

A high-capacity convolutional neural networks (CNNs) can be applied to bottom-up region proposals for localizing and segmenting of the objects and

When training data which is labeled is very less, monitored pre-training of an auxiliary task with large amount of data, then fine-tuning of the network of the target task with very less data, gives boost in the performance.

Since region proposals are combined with CNNs, in this paper, the method is called R-CNN: Regions with CNN features.

The results in this paper were obtained by using a combination of classical tools from computer vision and deep learning and conclude that they are partners.

III. CONCLUSION

From studying the research papers done in the field of machine learning, CNN, deep learning to recognise materials and classify images, we have gained insights about the various techniques that can be implemented. We have learnt about the techniques that can be incorporated to build a model for recognising the materials based on the images taken of them. This literature survey has enriched our knowledge on the studies and developments in the area thus giving us a clear picture about implementation.

IV. REFERENCE

- [1] Swanson, Kirk, et al. "Deep learning for automated classification and characterization of amorphous materials." *Soft Matter* (2020).
- [2] Saragadam, Vishwanath, and Aswin C. Sankaranarayanan. "Programmable Spectrometry--Per-pixel Classification of Materials using Learned Spectral Filters." *arXiv preprint arXiv:1905.04815* (2019).
- [3] Erickson, Zackory, et al. "Classification of household materials via spectroscopy." *IEEE Robotics and Automation Letters* 4.2 (2019): 700-707.
- [4] Li, Shutao, et al. "Deep learning for hyperspectral image classification: An overview." *IEEE Transactions on Geoscience and Remote Sensing* 57.9 (2019): 6690-6709.
- [5] Xue, Jia, et al. "Differential angular imaging for material recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
- [6] Schwartz, Gabriel, and Ko Nishino. "Recognizing material properties from images." *IEEE transactions on pattern analysis and machine intelligence* (2019).
- [7] Su, Shuochen, et al. "Material classification using raw time-of-flight measurements." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
- [8] Schmidt, Philipp, and Roland W. Fleming. "Identifying shape transformations from photographs of real objects." *PLoS one* 13.8 (2018).
- [9] Sticlaru, Anca. "Material Classification using Neural Networks." *arXiv preprint arXiv:1710.06854* (2017).
- [10] Schwartz, Gabriel, and Ko Nishino. "Material recognition from local appearance in global context." *arXiv preprint arXiv:1611.09394* (2016).
- [11] Touvron, Hugo, et al. "Fixing the train-test resolution discrepancy." *Advances in Neural Information Processing Systems*. 2019.
- [12] Kalliatakis, Grigorios, et al. "Material Classification in the Wild: Do Synthesized Training Data Generalise Better than Real-World Training Data?." *arXiv preprint arXiv:1711.03874* (2017).
- [13] Mangal, Ankita, and Elizabeth A. Holm. "A comparative study of feature selection methods for stress hotspot classification in materials." *Integrating Materials and Manufacturing Innovation* 7.3 (2018): 87-95.
- [14] Schwartz, Gabriel, and Ko Nishino. "Integrating Local Material Recognition with Large-Scale Perceptual Attribute Discovery." *arXiv preprint arXiv:1604.01345* (2016).
- [15] Bell, Sean, et al. "Material recognition in the wild with the materials in context database." *Proceedings of the*



IEEE conference on computer vision and pattern recognition. 2015.

- [16] Wang, Oliver, et al. "Material classification using BRDF slices." *2009 IEEE Conference on Computer Vision and Pattern Recognition.*
- [17] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [18] Zhou, Bolei, Vignesh Jagadeesh, and Robinson Piramuthu. "Conceptlearner: Discovering visual concepts from weakly labeled image collections." *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015.
- [19] Haldekar, Mandar, Ashwinkumar Ganesan, and Tim Oates. "Identifying spatial relations in images using convolutional neural networks." *2017 International Joint Conference on Neural Networks (IJCNN).* IEEE, 2017.
- [20] Barron, Jonathan T., and Jitendra Malik. "Shape, albedo, and illumination from a single image of an unknown object." *2012 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 2012.
- [21] Hu, Diane, Liefeng Bo, and Xiaofeng Ren. "Toward Robust Material Recognition for Everyday Objects." *BMVC.* Vol. 2. 2011.
- [22] Schwartz, Gabriel, and Ko Nishino. "Visual material traits: Recognizing per-pixel material context." *Proceedings of the IEEE International Conference on Computer Vision Workshops.* 2013.
- [23] Donahue, Jeff, et al. "Decaf: A deep convolutional activation feature for generic visual recognition." *International conference on machine learning.* 2014.
- [24] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014.