IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, May 2024

Object Detection using Convolutional Neural Network

Dr. B. V. Pranay Kumar[1], P. Rahul[2], S. Avinash[3], K. Shyamsundar[4], B. Sandhya[5], B. Sandhya[6]
Associate Professor, Department of Computer Science & Engineering[1]
UG Students, Department of Computer Science and Engineering[2,3,4,5,6]
Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

Abstract: *During the last years, a noticeable growth is observed in the field of computer vision research. In computer vision, object detection is a task of classifying and localizing the objects in order to detect the same. The widely used object detection applications are human– computer interaction, video surveillance, satellite imagery, transport system, and activity recognition. In the wider family of deep learning architectures, convolutional neural network (CNN) made up with set of neural network layers is used for visual imagery. Deep CNN architectures exhibit impressive results for detection of objects in digital image. This paper represents a comprehensive review of the recent development in object detection using convolutional neural networks.*

Keywords: computer vision, convolutional neural networks, image classification, object detection, transferlearning

I. INTRODUCTION

Object detection is important in computer vision systems. It can be used for many applications like video surveillance [1], medical imaging [2], and robot navigation [3]. Many algorithms can be used for this task like background subtraction, temporal differencing, optical flow, Kalman filtering, support vector machine, and contour matching [4]. Aside from the said algorithms, the newest method used for object detection is called convolutional neural networks (CNN). Breakthroughs in image classification started when
Alex [5] won the 2012 ImageNet competition using deep convolutional neural networks. They trained a deep CNN to classify 1.2 million high resolution images in the ImageNet [6] contest that has 1000 categories. They achieved more accurate prediction than the previous state of the art models. From this,many researchers became interested in finding a novel way todevelop an efficient deep convolutional neural networks.
There are some recent approaches for object detection, in paper [7], the authors developed a low shot transfer detector using a flexible deep architecture and a regularized transfer learning framework to address object detection using few training data. Another paper in [8] proposed a region selection network and a gatting network for object detection. The region selection network serves as guidance on where to select regions to learn the features from. On the other hand, gatting network serves as local feature selector that transforms feature maps. In [9] used convolutional neural 978-1-5386-5457- 6/18/$31.00 ©2018 IEEE networks for visual target tracking. They created an ad-hoc large dataset with positive and negative examples of framed objects from ImageNet databaseand selected the most promising patch using AlexNet [5] architecture. Another method used for object detection is active learning [10]. This is a class of algorithms that searches for the most informative samples to include in a training dataset that is effective in image classification. Lastly, in
[11] used artificial neural networks to detect objects by shape and color pattern recognition. This paper is organized as follows; Section II is a brief introduction about convolutional neural networks. Section III is discusses the concept of transfer learning. Section IV discusses the different TensorFlow models used for object detection. Section V shows the experiment set-up as well as the information about the dataset. Section VI is the results discussion.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, May 2024

II. CONVOLUTIONAL NEURALNETWORKS

Convolutional neural network (CNN) is a class of deep, feed-forward artificial neural network that has been utilized to produce an accurate performance in computer vision tasks, such as image classification and detection [5]. CNNs are like traditional neural network, but with deeper layers. It has weights, biases and outputs through a nonlinear activation. The neurons of the CNN are arranged in a volumetric fashion such as, height, width and depth.

Fig. 1 shows the CNN architecture, it is composed of convolutional layer, pooling layer and fully connected layer. Convolutional layer and pooling layer are typically alternated and the depth of each filter increases from left to right while the output size (height and width) are decreasing. The fully connected layer is the last stage which is similar to the last layer of the conventional neural networks.

The input is an image that will hold pixel values. It has three dimensions such as width, height and depth (RGB channels) example is [50 x 50 x 3] [13]. The convolutional layer will compute the output of neurons that are connected to local regions in the input. The layer's parameters are composed of a set of learnable filters (or kernels), which convolved across the width and height of the input volume extending through its depth, computing the dot product between the entries of the input and the filter. This produces a 2-dimensional activation map of that filter and as a result, the network learns filters that trigger when it detects some particular type of feature at some spatial position in the input. The function called Rectified Linear Unit (ReLU) layer will perform elementwise activation function.

This function is zero for negative values and grows linearly for positive values. This will not affect the volume size. The pooling layer outputs the maximum activation in a region. This down samples the spatial dimensions such as width and height. The output layer is the fully connected layer which is similar to the final layer of the neural network. This layer used commonly used softmax activation to output probability distributions over the number of output classes.

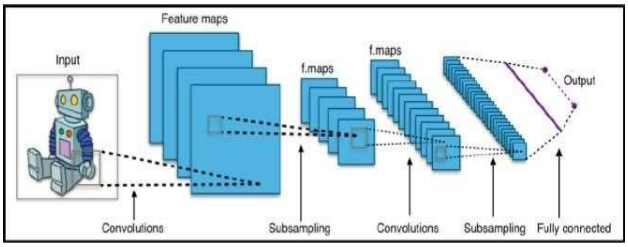


Fig. CNN Architecture

III. MODULES

Data Preprocessing Module:

- Purpose: Responsible for preparing the input dataset fortraining and evaluation.
- Tasks**:** Data loading: Reads annotated images and corresponding ground truth bounding boxes from the dataset.
- Data augmentation: Applies augmentation techniques such as rotation, flipping, scaling, and translation to increase datasetdiversity and model robustness.
- Data normalization: Normalizes pixel values of images to ensure consistency and convergence during training.
- Implementation: Utilizes Python libraries such as OpenCV and NumPy for image processing and manipulation.

Convolutional Neural Network (CNN) Module:

- Purpose: Implements the core object detection model based on CNN architecture.
- Tasks: Feature extraction: Learns hierarchical features from input images using multiple convolutional layers.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, May 2024

- Object localization: Predicts bounding boxes and class probabilities for objects present in the input image.
- Training and inference: Conducts both model training and inference phases, adjusting model parameters based on input dataand optimizing predictions for unseen data.
- Implementation: Utilizes deep learning frameworks such as TensorFlow or PyTorch to define and train the CNN model architecture.

Evaluation Module:

- Purpose: Evaluates the performance of the trained model on validation and test datasets.
- Tasks: Metric calculation: Computes evaluation metrics such asmean Average Precision (mAP), precision, recall, and F1-score to quantify model performance.
- Visualization: Generates visualizations of detection results, including bounding boxes overlaid on input images and associated confidence scores.
- Implementation: Utilizes Python libraries such as scikit-learn and Matplotlib for metric calculation and result visualization.

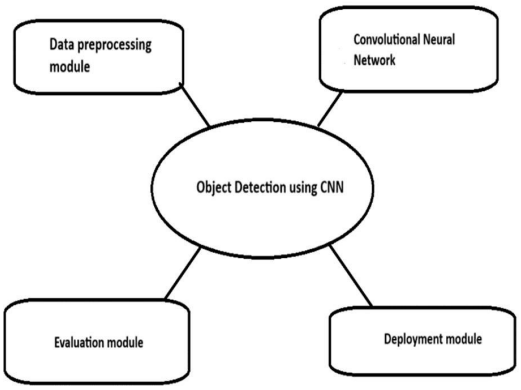
Deployment Module:

- Purpose: Facilitates the deployment of the trained objectdetection model for real-world applications.
- Tasks: Integration: Integrates the trained model into existing software systems or applications for object detection tasks. Optimization: Optimizes the model for deployment on targethardware platforms, such as CPUs, GPUs, or edge devices.

IV. TRANSFER LEARNING

Transfer learning is a powerful deep learning technique in which pre-trained models can be used for feature extraction and fine tuning. This technique can be used in image classification like vehicle classification [14], object detection and segmentation. The advantage of using this technique is it saves time to train the network from the start and less data is needed to get good results. The training can also be done using a central processing unit (CPU) even without the computational power of graphics processing unit (GPU). Instead of creating new model from scratch, pre-trained models can be used. These models are trained from the large image database like ImageNet [6] and COCO [15] dataset. Some of these models are AlexNet [5], VGG16, VGG19 [16], ResNet50 [17], InceptionV2, InceptionV3 [18], Xception [19], DenseNet [20] and MobileNet [21]. This paper focuses on the two models only, InceptionV2 and MobileNet.

SYSTEM ARCHITECTURE



The system architecture for the "Object Detection using CNN with Python" project comprises four interconnected modules.The Data Preprocessing Module prepares the dataset byloading annotated images, applying augmentation techniques for diversity, and normalizing pixel values. The Convolutional Neural Network (CNN) Module

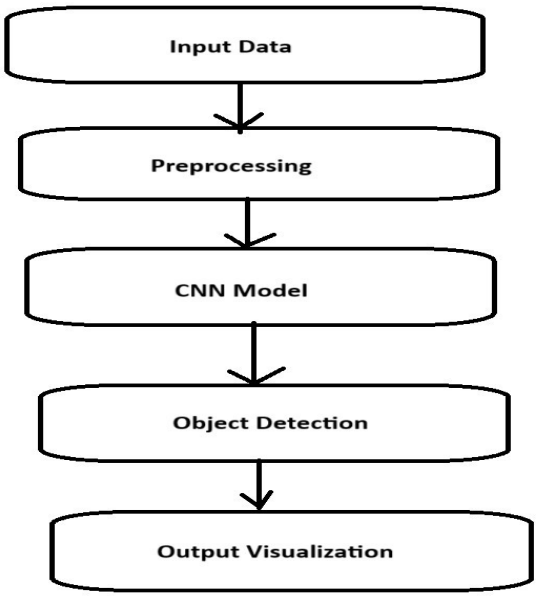
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, May 2024

implements the core object detection model, learning features, predicting bounding boxes, and handling training/inference phases. The Evaluation Module assesses model performance by computing metrics like mean Average Precision (mAP) and generating visualizations of detection results. Lastly, the Deployment Module facilitates the integration of the trained model into production systems, optimizing it for deployment on various hardware platforms and developing APIs/interfaces for seamless interaction. This modular architecture enables the development of an efficient and scalable object detection system using CNN with Python, from data preprocessing to practical deployment in real-world applications.

V. SOFTWARE DESIGN

Data Flow Diagram

A Data Flow Diagram (DFD) in UML (Unified Modeling Language) is a graphical representation of the flow of data through a system. It is used to model the processes and data involved in a system or application, and to illustrate the data flow between different components of the system. DFDs can be used to model systems at various levels of detail, from high-level overviews to detailed specifications. They can help identify data dependencies and interactions between different components of a system and can be used to communicate system requirements and specifications to stakeholders and development teams. It may be used as a communication tool between a system analyst and any person who plays a part in the order that acts as a starting point for redesigning a system. The DFD is also called as a data flow graph or bubble chart.



VI. TENSORFLOW OBJECT DETECTION MODELS

TensorFlow is an open-source software library for high performance numerical computation [22]. The following TensorFlow models are used to detect quadrotor unmanned aerial vehicle (UAV) [23] and person in an image and video.

SSD with MobileNetV1

A Single Shot Multi-Box Detector (SSD) is an object detection approach in images using a deep neural network [24]. It produces bounding box and class scores of the detected object at greater speed than previous approach like You Only Look Once (YOLO) [25]. On the other hand, MobileNet is a deep learning model that can be applied to various recognition tasks like object detection, finegrain classification, landmark recognition, and face attributes [21]. The advantages of MobileNet are accurate, fast, small and easy to tune. It is based on a simplified architecture that uses depthwise separable convolutions to build light weight deep neural networks [21]. SSD models that used

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, May 2024

MobileNet in object detection arelightweight so that in can be deployed in mobile devices and can be run in real-time [26].

Faster-RCNN with InceptionV2

Faster-RCNN is a single, unified network for object detection. It uses region proposal network (RPN) module that serves as the attention mechanism [27]. It tells the unified network where to look. On the other hand, Inception is composed of efficient inception module with 22 layers and no fully connected layers [18]. The main advantage of this model is the improved utilization of the computing resources inside the network. Inception module is a network within a networkand modules are stack on top of each other. It has 5 million parameters that are 12 times less than AlexNet model. Faster- RCNN together with InceptionV2 is computationally intensive but produces more accurate results in object detection.

VII. EXISTING SYSTEM

Often rely on handcrafted features and shallow classifiers (e.g., Haar cascades, HOG features). Limited in handling complex scenarios with occlusions, scale variations, andcluttered backgrounds.
Less efficient in terms of both accuracy and computational speed compared to deep learning-based approaches. Deep learning-based systems generally outperform traditional methods in terms of accuracy and robustness.
CNN-based approaches offer faster inference times, making them suitable for real-time applications .Transfer learning enables efficient training on limited datasets, reducing the need for large-scale annotated data.
Existing systems may vary in terms of programming languages, frameworks, and hardware dependencies.
Some systems may offer pre-trained models or cloud-based APIs for easy integration.
Considerations such as model size, memory footprint, and computational requirements influence deployment choices.
Deep learning-based systems tend to achieve higher accuracy compared to traditional methods, especially in challenging scenarios.
CNN-based approaches offer faster inference times, making them suitable for real-time applications. Ease of Use Traditional methods may be simpler to implement and deploy, but deep learning-based approaches offer better performance with proper training and optimization. Resource Requirements: Deep learning models often require more computational resources (e.g., GPUs) for training and inference compared to traditional methods. Generalization: Transfer learning allows deep learning models togeneralize well even with limited annotated data, whereas traditional methods may struggle with diverse datasets.

Disadvantages of Existing System:

- Limited Accuracy
- High Computational Cost
- Lack of Flexibility
- Slow Inference Speed
- Lack of Generalization
- Complexity of Implementation
- Sensitivity to Hyperparameters
- Limited Robustness to Variations
- Ethical and Privacy Concerns
- Maintenance and Updates

VIII. PROPOSED SYSTEM

Utilizes Convolutional Neural Networks (CNNs) for object detection.May employ architectures such as YOLO (You Only Look Once), SSD (Single Shot Multibox Detector), or Faster R- CNN. Allows for end-to-end learning of object features and spatial information. Data preprocessing includes augmentation techniques to increase dataset diversity. Utilizes transfer learning with pre-trained CNN models for faster convergence and better

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, May 2024

generalization. Fine-tunes model parameters on the specific object detection task. Optimizes hyperparameters such as learning rate and batch size for improved performance. Evaluates model performance using metrics like mean Average Precision
,precision, recall, and F1-score. Conducts validation on a separate dataset to assess generalization ability. Visualizes detection results with bounding boxes and class labels overlaid on images.
Developed using Python programming language with libraries/frameworks such as TensorFlow, Keras, and OpenCV. Modular design allows for easy integration into other applications or systems. Offers flexibility in deployment, supporting various hardware platforms.

Advantages of Proposed System:

- Established Techniques
- Compatibility with Limited Hardware
- Familiarity and Accessibility
- Robustness to Small Datasets
- Versatility in Application Domains
- Incremental Improvement Potential
- Adaptability to Resource Constraints
- Interpretability and Explainability
- Cost-effectiveness in Certain Scenarios
- Legacy System Integration

IX. EXPERIMENT SETUP

To implement the object detection using CNN, TensorFlow Object detection API [28] was used. This is an open source framework for constructing, training and deploying object detection models. The dataset used for this research was limited to a person and quadrotor only. Training images were extracted from personal video while testing images were from the internet. Table I shows the number of images used for training and testing. As a rule of thumb, 80% of the images were utilized for training and 20% for testing. The images were labeled manually using LabelImg, a graphical image annotation tool [29]. This creates XML files in PASCAL VOC format [30].

X. RESULTS

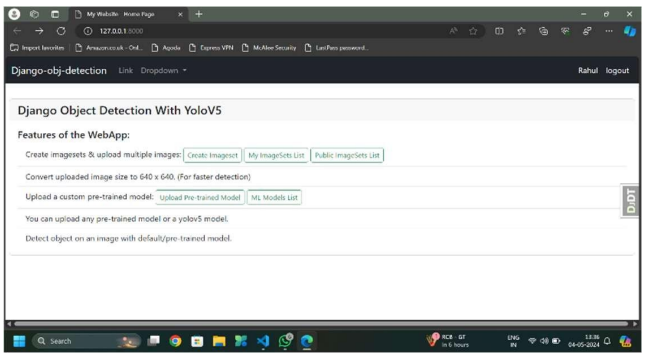


Figure1:A Home page for Object detection

Fig . This home page for object detection. Here ,They upload the image in the create imageset and by selecting the model in it, the object is detected. The Detected Object Images is saved in the My Imagesets List Ssection and we train the Models aas we need to Detect any kind of object

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, May 2024

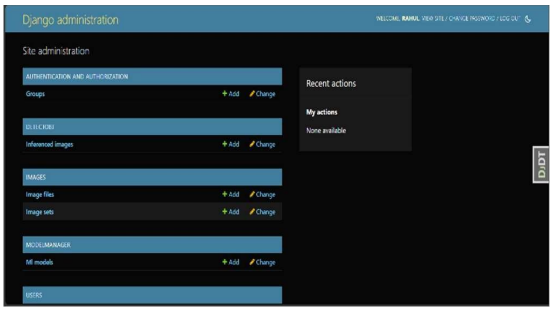


Figure.2. Django admin page for Object detection

In fig 2 In this the Admin Page has the details of the saved objects that they detected with this the admin can add the users and also at the same delete them.

Figure 3. Uploading image and detecting the multiple objectsin that image a time

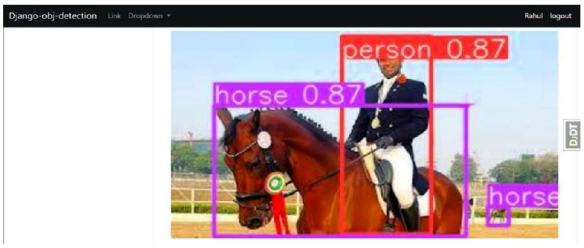


Figure 4. Uploading image and detecting the multiple objectsin that image a time

Figure 5. Uploading image and detecting the multiple objects in that image a time

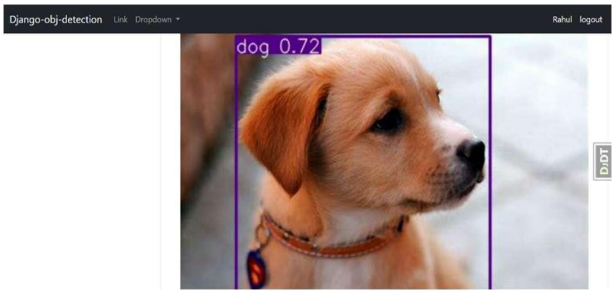


Figure 6. Uploading image and detecting single object at a time

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, May 2024

XI. CONCLUSION

The object detection capability of the two state of the art models in CNN was successfully demonstrated. It shows that the SSD with MobileNetV1 has high speed detection but low accuracy compared with Faster-RCNN with InceptionV2 that has low speed but more accurate. Base on the results of the experiments, there's a trade-off between accuracy and speed.

If we want fast detection capability especially in realtime applications, use SSD with MobileNetV1. If high accurate detection capability is desired, use Faster-RCNN with InceptionV2. For future research, the two models will be implemented as vision system of bomb disposal robot to detect improvised explosive devices (IED's).

ACKNOWLEDGMENT

The author would like to thank Department of Science and Technology – Engineering Research and Development for Technology (DOST-ERDT) and De La Salle University for the financial support while doing the study.

REFERENCES

[1]V. Gajjar, A. Gurnani and Y. Khandhediya, "Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach," in 2017 IEEE International Conference on Computer Vision Workshops, 2017.

[2]M. Adel, A. Moussaoui, M. Rasigni, S. Bourennane and L. Hamami, "Statistical-Based Tracking Technique for Linear Structures Detection: Application to Vessel Segmentation in Medical Images," IEEE Signal Processing Letters, vol. 17, no. 6, pp. 555-558, June 2010.

[3]X.-T. Truong, V. N. Yoong and T.-D. Ngo, "RGB-D and Laser Data Fusion-based Human Detection and Tracking for Socially Aware Robot Navigation Framework," in IEEE Conference on Robotics and Biomimetics, Zhuhai, China, 2015.

[4]H. S. Parekh, D. G. Thakore and U. K. Jaliya, "A Survey on Object Detection and Tracking Methods," International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, no. 2, pp. 2970-2978, February 2014.

[5]A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in neural information processing systems, 2012.

[6]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009.

[7]H. Chen, Y. Wang, G. Wang and Y. Qiao, "LSTD: A Low-Shot Transfer Detector for Object Detection," arXiv:1803.01529v1, 2018.

[8]H. Xu, X. Lv, X. Wang, Z. Ren and R. Chellappa, "Deep Regionlets for Object Detection," arXiv:1712.02408v1, December 2017.

[9]"Deep learning to frame objects for visual target tracking," Engineering Applications of Artificial Intelligence, vol. 65, pp. 406420, October 2017.

[10]C.-C. Kao, P. Sen, T.-Y. Lee and M.-Y. Liu, "Localization- Aware Active Learning for Object Detection," arXiv:1801.05124v1, January 2018.

[11]J. P. N. Cruz, M. L. Dimaala, L. G. L. Francisco, E. J. S. Franco, A. A. Bandala and E. P. Dadios, "Object Recognition and Detection by Shape and Color Pattern Recognition Utilizing Artificial Neural Networks," in 2013 International Conference of Information and Communication Technology (ICoICT), Bandung, Indonesia, 2013.

[12]A. Gulli and S. Pal, Deep Learning with Keras, Birmingham: Packt, 2017.

[13]A. Karpathy, October 2017. [Online]. Available: http://cs231n.github.io/convolutional-networks/.

[14]R. L. Galvez, N. N. F. Giron, M. K. Cabatuan and E. P. Dadios, "Vehicle Classification Using Transfer Learning in Convolutional Neural Networks," in 2017 2nd Advanced Research in Electrical and Electronic Engineering Technology (ARIEET), Jakarta, Indonesia, 2017.

[15]T.-Y. Lin, J. Hays, M. Maire, P. Perona, S. Belongie, D. Ramanan, L. Bourdev, L. Zitnick, R. Girshick and P. Dollar, "Microsoft COCO: Common Objects in Context," arXiv:1405.0312v3, pp. 1-15, February 2015.