Deep Learning for X-ray Image to Text Generation

Mahima Chaddha, Sneha Kashid, Snehal Bhosale, Prof. Radha Deoghare

Department of Information Technology, PCET's Nutan Maharashtra Institute of
Engineering and Technology, Talegaon Dabhade, Maharashtra, India

How to cite this paper: Mahima Chaddha | Sneha Kashid | Snehal Bhosale | Prof. Radha Deoghare "Deep Learning for X-ray Image to Text Generation" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-3, April 2019, pp.1679-1682, URL: https://www.ijtsrd.com/papers/ijtsrd23168.pdf

IJTSRD23168

ABSTRACT

Motivated by the recent success of supervised and weakly supervised common object discovery, in this work we move forward one step further to tackle common object discovery in a fully unsupervised way. Mainly, object co-localization aims at simultaneously localizing the objects of the same class across a group of images. Traditional object localization/ detection usually trains the specific object detectors which require bounding box annotations of object instances, or at least image-level labels to indicate the presence/absence of objects in an image. Given a collection of images without any annotations, our proposed fully unsupervised method is to simultaneously discover images that contain common objects and also localize common objects in corresponding images.

It has been long envisioned that the machines one day will understand the visual world at a human level of intelligence. Now we can build very deep convolutional neural networks (CNNs) and achieve an impressively low error rate for tasks like large-scale image classification. However, in tasks like image classification, the content of an image is usually simple, containing a predominant object to be classified. The situation could be much more challenging when we want computers to understand complex scenes. Image captioning is one such task. In these tasks, we have to train a model to predict the category of a given x-ray image is to first annotate each x-ray image in a training set with a label from the predefined set of categories. Through such fully supervised training, the computer learns how to classify an x-ray image and convert into text.

KEYWORDS: *object detection, object tracking, object identification, edge detection, convolutional neural networks (CNNs).*

I. INTRODUCTION

Whenever shown the image, our brain instantly recognizes a objects contained in it. On the other hand, it takes a lot of time and the training data for a machine to identify these objects. But with the recent advances in hardware and the deep learning, this computer vision field has become a whole lot easier and more intuitive. We are constantly in the search of methods to have a 'detection' or 'recognition' system as powerful as the human being.

Weakly supervised Object localization(WSOL), has drawn much attention recently. It aims at localizing common objects across images using the annotations to indicate the presence/absence of the objects of interest. In this project we focus on simultaneously discovering and localizing common objects in real world images, which shares the same type of output as WSOL, but does requires the annotation of presence/absence of objects. in addition, we tackle this problem in more challenging scenario where,

1. Multiple common object classes are contained in the given collection of images, which means this is totally unsupervised problem
2. multiple objects or even no objects is contained in -some of the images.

The project aims to incorporate the state-of-the-art technique for the object detection with the goal of achieving high the accuracy with a real-time performance. The major challenge in many of the object detection systems is that the dependency on other computer vision techniques for helping the deep learning based approach, which leads to slow and non-optimal performance. In the project, we use a completely deep learning based approach to solve the problem of the object detection in an end-to-end fashion.

The situation could be much more challenging when we want computers to understand complex scenes. Image captioning is one such task. In these tasks, we have to train a model to predict the category of a given image is to first annotate each image in a training set with a label from a predefined set of categories.

II. LITERATURE REVIEW

1. **Paper name**: Object Detection Using Image Processing
   Author: Fares Jalled, ´ Moscow Institute of Physics & Technology

In this paper ,they have develop an Open CV-Python code using Haar Cascade algorithm for object and face detection. Currently, UAVs are used for detecting and attacking the infiltrated ground targets. The main drawback for this type of UAVs is that sometimes the object are not properly detected, which thereby causes the object to hit the UAV. This project aims to avoid such unwanted collisions and damages of UAV. UAV is also used for surveillance that uses Voila-jones algorithm to detect and track humans. This algorithm uses cascade object detector function and vision.

2. **Paper name:** Edge Preserving and Multi-Scale Contextual Neural Network for Salient Object Detection.
   Author: Xiang Wang , Huimin Ma , Member IEEE, Xiaozhi Chen, and Shaodi You.

In this paper, we propose a novel edge preserving and multi-scale contextual neural network for salient object detection. The proposed framework is aiming to address two limits of the existing CNN based methods. First, region-based CNN methods lack sufficient context to accurately locate salient object since they deal with each region independently. Second, pixel-based CNN methods suffer from blurry boundaries due to the presence of convolutional and pooling layers. Motivated by these, we first propose an end-to-end edge-preserved neural network based on Fast R-CNN framework (named RegionNet) to efficiently generate saliency map with sharp object boundaries. The proposed framework achieves both clear detection boundary and multiscale contextual robustness simultaneously for the first time, and thus achieves an optimized performance. Experiments on six RGB and two RGB-D benchmark datasets demonstrate that the proposed method achieves state-of-the-art performance.

3. **paper name:** 3D Object Proposals using Stereo Imagery for Accurate Object Class Detection.
   Author: Xiaozhi Chen∗ , Kaustav Kundu∗ , Yukun Zhu, Huimin Ma, Sanja Fidler and Raquel Urtasun.

In this paper, a novel 3D object detection approach is implemented that exploits stereo imagery and contextual information specific to the domain of autonomous driving. We propose a 3D object proposal method that goes beyond 2D bounding boxes and is capable of generating highquality 3D bounding box proposals. We make use of the 3D information estimated from a stereo camera pair by placing 3D candidate boxes on the ground plane and scoring them via 3D point cloud features. In particular, our scoring function encodes several depth informed features such as point densities inside a candidate box, free space, visibility, as well as object size priors and height above the ground plane. The inference process is very efficient as all the features can be computed in constant time via 3D integral images

4. **Paper name:**Scalable Object Detection using Deep Neural Networks
   Author: Christian Szegedy, Dumitru Erhan, Alexander Toshkov Toshev

In this paper, a Deep convolutional neural networks have recently achieved state-of-the-art performance on a number of image recognition benchmarks, including the Image Net Large-Scale Visual Recognition Challenge (ILSVRC-2012). The winning model on the localization sub-task was a network that predicts a single bounding box and a confidence score for each object category in the image. Such a model captures the whole-image context around the objects but cannot handle multiple instances of the same object in the image without naively replicating the number of outputs for each instance. In this work, we propose a saliency-inspired neural network model for detection, which predicts a set of class-agnostic bounding boxes along with a single score for each box, corresponding to its likelihood of containing any object of interest. The model naturally handles a variable number of instances for each class and allows for cross-class generalization at the highest levels of the network. We are able to obtain competitive recognition performance on VOC2007 and ILSVRC2012, while using only the top few predicted locations in each image and a small number of neural network evaluations.

5. **Paper name:** Rich feature hierarchies for accurate object detection and semantic segmentation
   Author: Ross Girshick1 Jeff Donahue1,2 Trevor Darrell1,2 Jitendra Malik1 1UC Berkeley and 2 ICSI

In this paper, we propose a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 30% relative to the previous best result on VOC 2012—achieving a mAP of 53.3%. Our approach combines two key insights: (1) one can apply high-capacity convolutional neural networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data is scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance boost. Since we combine region proposals with CNNs, we call our method R-CNN: Regions with CNN features.

6. **Paper name:** Image-Text Surgery: Efficient Concept Learning in Image Captioning by Generating Pseudopairs
   Author: Kun Fu , Jin Li, Junqi Jin, and Changshui Zhang, Fellow, IEEE

In this paper, they used semantic structure of image and text to efficiently learn novel concepts. We noticed that both images and sentences consist of several semantic meaningful components that can be shared across image-sentence pairs. For example, “a zebra/giraffe in a green grassy field” shares the context “in a green grassy field.” Combining zebra or giraffe with the context is both logically correct. Such a semantic structure enables a more efficient way to learn novel concepts. Suppose the system has learned the concept of giraffe but has never seen a zebra, it can learn to describe a zebra in a field, just by recognizing zebra and knowing the fact that a zebra can be in a grassy field like a giraffe.The image and sentence are thus decoupled—the required data sources of novel concepts consist of: an independent image base providing visual information and an independent knowledge base providing logic information.

7. **Paper name:**Predicting Visual Features from Text for Image and Video Caption Retrieval.
   Author: Jianfeng Dong, Xirong Li, and Cees G. M. Snoek

In this paper, they strives to find amidst a set of sentences the one best describing the content of a given image or video. Different from existing works, which rely on a joint subspace for their image and video caption retrieval, we propose to do so in a visual space exclusively. Apart from this conceptual novelty, we contribute Word2VisualVec, a deep neural network architecture that learns to predict a visual feature representation from textual input. Example captions are encoded into a textual embedding based on multi-scale sentence vectorization and further transferred into a deep visual feature of choice via a simple multi-layer perceptron.

III. EXISTING SYSTEM

Localizing and detecting objects in images are among the most widely studied computer vision problems. They are quite challenging due to intra-class variation, inter-class diversity, and noisy annotations, especially in wild images. Thus, a large body of fully/strongly annotated data is crucial to train detectors to achieve satisfactory performance. Early approaches to image captioning can be roughly divided into two families. The first one is based on template matching. These approaches start from detecting objects, actions, scenes, and attributes in images and then fill them into a hand-designed and rigid sentence template. The captions generated by these approaches are not always fluent and expressive. The second family is grounded on retrieval based approaches, which first select a set of the visually similar images from a large database and then transfer the captions of retrieved images to fit the query image. There is little flexibility to modify words based on the content of the query image, since they directly rely on captions of training images and cannot generate new captions.

IV. PROPOSED SYSTEM

The main paradigm of these tasks are similar: the inputs are usually xray images with incomplete labels or sometimes even without any supervision information, then the key step is to discover the most frequently occurring pattern by methods such as local feature matching, sub-graph mining, etc.

In these tasks, we have to train a model to predict the category of a given image is to first annotate each image in a training set with a label from a predefined set of categories. Through such fully supervised training, the computer learns how to classify an image by using CNN and RNN.

ADVANTAGES OF PROPOSED SYSTEM:

Our proposed framework can also be easily applied in the problem of image/instance retrieval.

Deep neural networks can potentially address both of these issues by generating fluent and expressive captions, which can also generalize beyond those in the train set.

These automatic metrics can be computed efficiently.

They can greatly speed up the development of image captioning algorithms. However, all of these automatic metrics are known to only roughly correlate with human judgment

V. SYSTEM ARCHITECTURE

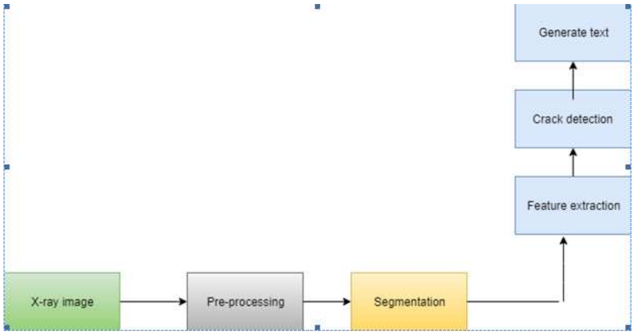

Fig.: System Architecture

VI. CONCLUSION AND FUTURE WORK

We propose a framework for common object discovery and localization in wild images. Like most previous methods which are based on the assumption that there is only one object contained in each positive image. Inspired by min-cut/max-flow algorithms. We can classify and detect the object by using neural network correctly.

We have studied detection techniques into various categories, here, we also discuss the related issues, to the object detection technique. This project gives valuable insight into this important research topic and encourages the new research in the area of moving object detection as well as in the field of computer vision. In image detection approach, various estimating methods are used to find corresponding region to target the defect.

Motivation -

Object recognition is one of the fundamental tasks in computer vision. It is the process of finding or identifying instances of objects (for example faces, dogs or buildings) in digital images or videos. Object recognition methods frequently use extracted features and learning algorithms to recognise instances of an object or images belonging to an object category. Objects in the images are detected and relation in between the objects are identified. Every object or object class has its own particular features that characterise themselves and differentiate them from the rest, helping in the recognition of the same or similar objects in other images or videos. Object recognition is applied in many areas of computer vision, including image retrieval, security, surveillance.

VII. REFERENCES

[1]. Bhavin V. Kakani, Divyang Gandhi, Sagar Jani, \ Improved OCR based Auto-matic Vehicle Number Plate Recognition using Features Trained Neural Net-work," International Conference on Communication and Network Technology, pp.1-6, IEEE-2017.

[2]. Anand Sumatilal Jain, Jayshree M. Kundargi, \ Automatic Number Plate Recognition Using Artificial Neural Network ,", International Research Journal of Engineering and Technology (IRJET), Vol.02, PP.1072-1078, 2015.

[3]. Pratiksha Jain ,Neha Chopra ,Vaishali gupta, , \ Automatic License Plate Recognition using OpenCV, ", International Journal of Computer Applications Technology and Research, Vol.3, pp. 756-761, 2014.

[4]. Utkarsh Dwivedi, Pranjal Rajput, Manish Kumar Sharma, \ License Plate Recognition System for Moving Vehicles Using Laplacian Edge Detector and Feature Extraction ,", International Research Journal of Engineering and Tech-nology (IRJET), Vol 4, pp.407-412, 2017.

[5]. Gajendra Sharma, \Performance Analysis of Vehicle Number Plate Recognition System Using Template Matching Techniques,", Journal of Information Tech-nology Software Engineering, Vol 8, pp.1-9, 2018.

[6]. Muhammad Tahir Qadri, Muhammad Asif, \Automatic Number Plate Recogni-tion System For Vehicle Identification Using Optical Character Recognition ,", International Conference on Education Technology and Computer, pp 335-338, IEEE-2009.

[7]. Chao-Ho Chen, Tsong-Yi Chen, Min-Tsung Wu, Tsann-Tay Tang, Wu-Chih Hu, \License Plate Recognition for Moving Vehicles Using a Moving Camera," International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp.497-500, IEEE-2013

[8]. Chuin-Mu Wang, Jian-Hong Liui, \ License Plate Recognition System,", Inter-national Conference on Fuzzy Systems and Knowledge Discovery, pp.1708-1710, 2015.

[9]. Abhishek Sharma, Amey Dharwadker, Thotreingam Kasar, \MobLP: A CC-based approach to vehicle license plate number segmentation from images ac-quired with a mobile phone camera," IEEE India Conference, pp.1-4, 2010.

[10]. Teik Koon Cheang, Yong Shean Chong, Haur Tay \Segmentation-free Vehicle License Plate Recognition using ConvNet-RNN,",International Workshop on Advanced Image Technology, pp. 1-5, 2017.