Deep Learning Applications in Computer Vision

Darshana Patel
VPMP Polytechnic, Gandhinagar, India

Abstract— Over the last few years deep learning methods have been shown to do better than previous traditional machine learning techniques in several fields, with computer vision being one of the most prominent cases. This review paper provides a brief overview of some of the most significant deep learning application used in computer vision problems. This paper has a description of deep learning applications in various computer vision tasks, such as object detection, face recognition, Image colorization, Image super-resolution, action and activity recognition, human pose estimation and Image Classification.

Key words: Deep Learning, Computer Vision, Object Detection, CNN, SVM, Face Recognition, Human Pose Estimation

I. INTRODUCTION

Applications in Computer Vision - A works that have done in deep learning methods to address works in computer vision, such as object detection, face recognition, image colorization, Image super-resolution, action and activity recognition, and human pose estimation.

A. Object Detection.

Object detection is the task of image classification with localization, although an image may contain multiple objects that require localization and classification[1]. Object detection is the process of detecting instances of semantic objects of a certain class (such as humans or birds) in digital images and video [2]. A common approach for object detection includes the creation of a large set of candidate windows that are in the outcome classified using CNN features[2]. For example, the schemes described in [3] employs selective search [4] to derive object proposals, extracts CNN features for each proposal, and then feeds the features to an Support Vector Machine(SVM) classifier to decide whether the windows include the object or not. A large number of works is based on the concept of Regions with CNN features proposed in [3].

B. Face Recognition.

Face recognition is one of the demandable computer vision applications with huge commercial interest as well. A variety of face recognition systems based on the extraction of handcrafted features have been proposed [5-8]; in such case, a feature extractor extracts features from an aligned face to obtain a low-dimensional representation, based on which a classifier makes predictions [2].

C. Image Colorization

Image colorization or neural colorization involves converting a grayscale image to a full color image. This task can be consideration of as a type of photo filter or transform that may not have an objective evaluation. Examples include colorizing old black and white photographs and movies.

Datasets often involve using existing photo datasets and creating grayscale versions of photos that models must learn to colorize [1].

D. Image Super-Resolution

Image super-resolution is the task of generating a new version of an image with a higher resolution and detail than the original image. Often models developed for image super-resolution can be used for image restoration and in painting as they solve related problems. Datasets often involve using existing photo datasets and creating down-scaled versions of photos for which models must learn to create super-resolution versions[1].

E. Action and Activity Recognition.

Human action and activity recognition is a research issue that has received a lot of attention from researchers [9,10]. Many works on human activity recognition based on deep learning techniques have been proposed in the literature in the last few years [11].

F. Human Pose Estimation.

The goal of human pose estimation is to determine the position of human joints from images, depth images, image sequences, or skeleton data as provided by motion capturing hardware [12]. Human pose estimation is a very challenging task having vast range of human appearances, difficult illumination, and cluttered background[2]. Before the era of deep learning, pose estimation was based on detection of body parts, for example, through pictorial structures [13]. Moving on to deep learning methods in human pose estimation, can group them into holistic and part-based methods, depending on the way the input images are processed.

G. Image Classification

Image classification involves assigning a label to an entire image or photograph. This term is also referred to as “object classification” and more generally as a “image recognition”.
Some examples of image classification include[1]:

- Labeling an x-ray as cancer or not (binary classification).
- Classifying a handwritten digit (multiclass classification).
- Assigning a name to a photograph of a face (multiclass classification).

A popular example of image classification used as a standard problem is the MNIST[14] dataset.

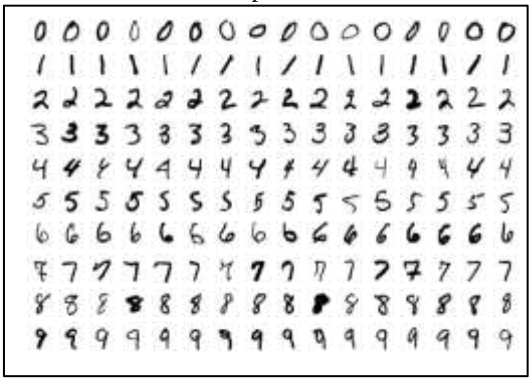

Fig. 1: Example of Hand written Digits from the MNIST Dataset

H. Image Synthesis

Image synthesis is task of generating targeted modifications of existing images or entirely new images.
This is a very broad area that is rapidly research and advancing [1].
It may include little bit modifications of image and video (i.e. image-to-image translations), such as:

- Changing the style of an object in a picture.
- Adding an object to a picture.
- Adding a face to a picture.

I. Deep Learning Datasets:

Deep learning datasets are collections of data specifically designed for use in deep learning models[15]. These datasets are typically large and complex, containing hundreds of thousands or even millions of data points1. Some of the most popular deep learning datasets include MNIST, MS-COCO, ImageNet, Open Images Dataset, VisualQA, and The Street View House Numbers (SVHN)[15]. These datasets are used to train deep learning models, which are algorithms that can learn from data and make predictions[15].

ACKNOWLEDGMENTS

I am sincerely thankful to the authors whose research papers helped me in making this work presentable.

REFERENCES

[1] Applications of Deep Learning for Computer Vision by Jason Brownlee on July 5, 2019 in Deep Learning for Computer Vision 39
[2] Deep Learning for Computer Vision: A Brief Review By Athanasios Voulodimos, 1,2 Nikolaos Doulamis,2 Anastasios Doulamis,2 and Eftychios Protopapadakis2, ,Hindawi-Article ID 7068349.
[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich fea- ture hierarchies for accurate object detection and semantic segmentation," in Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14), pp. 580– 587, Columbus, Ohio, USA, June 2014.
[4] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," Interna- tional Journal of Computer Vision, vol. 104, no. 2, pp. 154–171, 2013.
[5] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification," in Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13), pp. 3025– 3032, June 2013.
[6] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, "A practical trans- fer learning algorithm for face verification," in Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13), pp. 3208–3215, December 2013.
[7] T. Berg and P. N. Belhumeur, "Tom-vs-Pete classifiers and iden- tity-preserving alignment for face verification," in Proceedings of the 23rd British Machine Vision Conference (BMVC '12), pp. 1–11, September 2012.
[8] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: a joint formulation," in Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part III, vol. 7574 of Lecture Notes in Computer Science, pp. 566–579, Springer, Berlin, Germany, 2012.
[9] A. S. Voulodimos, D. I. Kosmopoulos, N. D. Doulamis, and T. A. Varvarigou, "A top-down event-driven approach for concurrent activity recognition," Multimedia Tools and Applications, vol. 69, no. 2, pp. 293–311, 2014.

[10] A. S. Voulodimos, N. D. Doulamis, D. I. Kosmopoulos, and T. A. Varvarigou, “Improving multi-camera activity recognition by employing neural network based readjustment,” Applied Arti- ficial Intelligence, vol. 26, no. 1-2, pp. 97–118, 2012.
[11] K. Makantasis, A. Doulamis, N. Doulamis, and K. Psychas, “Deep learning based human behavior recognition in indus- trial workflows,” in Proceedings of the 23rd IEEE International Conference on Image Processing, ICIP 2016, pp. 1609–1613, September 2016.
[12] A. Kitsikidis, K. Dimitropoulos, S. Douka, and N. Grammalidis, “Dance analysis using multiple kinect sensors,” in Proceedings of the 9th International Conference on Computer Vision Theory and Applications, VISAPP 2014, pp. 789–795, prt, January 2014.
[13] P. F. Felzenszwalb and D. P. Huttenlocher, “Pictorial struc- tures for object recognition,” International Journal of Computer Vision, vol. 61, no. 1, pp. 55–79, 2005.
[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Hafner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2323, 1998.
[15] https://www.ilovephd.com/free-deep-learning-data-sets/