wjert, 2022, Vol. 8, Issue 3, 161-165. *Review Article* ISSN 2454-695X

World Journal of Engineering Research and Technology

WJERT

www.wjert.org SJIF Impact Factor: 5.924

A STUDY ON THE METHODS OF IMAGE FEATURE EXTRACTION

***Sayeed Mohammed Shaik and [2]Dr. R. Kamalraj**

[1]MCA, MCA Department, Jain University – Bangalore.

[2]Associate Professor, MCA Department, Jain University – Bangalore.

Article Received on 10/01/2022 Article Revised on 30/01/2022 Article Accepted on 20/02/2022

***Corresponding Author**
Sayeed Mohammed Shaik
MCA, MCA Department,
Jain University – Bangalore.

ABSTRACT

Computer Vision tasks analyze the whole image. These tasks are computationally expensive. By extracting features from the image, we can reduce the computational expense. Using Deep Learning we can create Intelligent Models that can perform these Computer Vision tasks. We can tackle each task such as image classification and object detection with Models specifically trained to deal with those tasks. By combining said specifically trained Models we can achieve an aggregated output. Some of the techniques such as Instance Segmentation deal with the Computer Vision tasks in similar fashion. Features in deep learning can range from luminance of the pixels to high level pattern classification. Images are represented as an array of pixels, we can utilize Convolutional Neural Networks for the Image Processing removing the need to transform the array of pixels. Other Neural Network architectures such as U-net and Auto-Encoders can also be utilized in dealing with similar Computer Vision tasks. Many Models are trained to deal with specific tasks. This paper surveys some of the Models currently performing the Computer Vision tasks and Image processing.

KEYWORDS: Computer Vision, Image Processing, CNN.

INTRODUCTION

Biological visual systems evolved from a simple unicellular organism that follows light stimuli to humans that recognize a wide range of visible spectra and patterns. Not only do we interpret, we also navigate the world around us with interpreted information. We also represent information. through paintings, books and now digital media. Computers allowed us to capture and store visual information through images. It also allowed us to manipulate and create. With AI capabilities, we can design and build models that can analyze, extract and transform information from an image as required for the computational task. We need different methods and algorithms to optimize the process.

An image is represented as an array of values in the computer. This array of values contains color information for each pixel on the screen, mapped using Cartesian coordinates.

A single pixel may not convey the full information required for high-level tasks. The task of computer vision is to analyze and interpret information both as a whole and on an individual level, depending on the task at hand. Many libraries and technologies provide the necessary tools and algorithms for image processing. Simple Kernel operations can be applied on the Image which is generally referred as Filter using numpy. Which provides functionality of representing an Image as a numpy Array. Tasks such as Edge Detection can be achieved on a single Image by using filters such as Sobel etc. Other features can be extracted in a similar fashion by applying filters. While this workflow might seem fit for a single Image, Scaling it for fitting large datasets can become troublesome. Deep Learning Models can aid us in adapting the Operations for a larger scale.

Need for Feature Extraction

As established a single Image consists of a number of pixels respective to its dimensions i.e height and width. The amount of pixels can scale rapidly as the dimensions increase. It is computationally expensive to deal with every single pixel of the Image. By extracting features from the image we can reduce the number of parameters we have to deal with during computation of the Computer Vision and Image Processing tasks. Multiple methods exist to extract different features of a given Image. As discussed previously these operations are run across the whole image using a Kernel with size smaller than that of the given Image. Effects such as blurring and sharpening can be achieved by simple kernels. Kernels such as Sobel and Canny aid in detecting the edge in a given Image. The main task of Feature Extraction is to reduce the parameters of a given Image. By extracting different sets of features using different approaches best suitable for specific tasks and combining the extracted features we reduce the number of parameters.

With Deep Learning Models the features can range from configuration of pixels in a 3x3 matrix to high level patterns. These features are connected allowing us to perform operations at different levels/layers of the model. This not only helps us reduce the computational expense in the Image Processing and Computer Vision tasks but also aids in approaching output in categorical classification where simple rules seem not fit. With these models we can perform multiple calculations and produce more than one output depending on the requirements.

LITERATURE REVIEW

Multiple Models are currently being implemented across various domains for domain specific tasks. As the scope of this study we are going to have an overview of some of the implementations.

Mask R-CNN does Instance Segmentation on an Image or a Video. It outputs bounding boxes, masks, and classes for each instance of an object detected in the Image or Video. It does so, By proposing regions for the objects in the Image. Classifying the object. Refining the Bounding boxes. Finally classifying pixels by generating masks. These steps/operations are run across all the objects in the given Image or Video. Mask R-CNN is an extension of Faster R-CNN. Using a ConvNet Faster R-CNN extracts feature maps from the images. Bounding boxes are obtained by passing these feature maps through a Region Proposal Network in Faster R-CNN. The main difference between the Mask R-CNN and the Faster R-CNN is that, Mask R-CNN returns Object mask in addition to the bounding boxes. Mask R-CNN is utilized in multiple domains for Image Processing and Computer Vision tasks. Self Driving Vehicles, Segmentation of cells in Medical Imagery, Detecting and enhancing subjects in a Photo are some examples of where Mask R-CNN is currently being used. In simpler terms we can think about Mask R-CNN as separating the Image into multiple Images for each object. Mask R-CNN can run on a single Image or Sequence of Images.

Context Encoders: Feature Learning by Inpainting is an Unsupervised feature learning algorithm. Inpainting can be thought of as filling in the missing pixel values of a given Image by analyzing the and recognizing patterns to generate pixel values that fit their surroundings seamlessly. An example of Inpainting using a method from Skimage can be seen in Fig 1. As we can see in Fig 1 even though the missing parts are filled. It fails to establish proper shapes and patterns of the buildings in the Image.

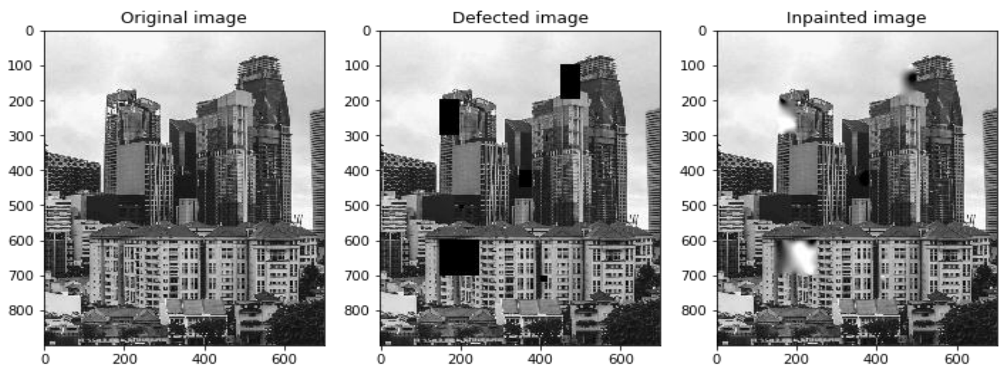

Fig. 1: Inpainting example.

The paper proposes a model that understands the whole image and proposes a plausible hypothesis for the missing parts. With the learned information the model can fill in the missing parts of the image with suitable patterns and shapes that make it seem like an original photo.

One Shot 3D Photography generates a 3D parallax effect output from a single image, either by uploading an Image or capturing the Image form a mobile. It is integrated into the Facebook app as of now and can only be used to upload images to Facebook. An explanation of the model approaches the task. A Layered depth image is generated from the given Image.A dense depth map is obtained by a proposed architecture called Tiefenrausch, which combines Efficient Block Structures in a U-net architecture. Further the process continues to generate Texture Atlas and Mesh.

Texture Atlas is obtained by the proposed methods, Depth Preprocessing, Occluded-Surface Hallucination which are run inside the neural net. Filter padding Maps the PConv Network to LDI texture atlas. A single atlas image is packed into a single texture using Macroblock padding. Mesh is generated by 2d triangulation of the extracted objects from the image. Vertices in the mesh are mapped as per the extracted depth map from the net.The model is trained for giving high performance in the mobile devices. The model is integrated into Facebook for convenient sharing of the output.

CONCLUSION

By extracting the Features of an Image we reduce the number of parameters/variables from the number of pixels in the Image to a smaller number. This can also aid in classification of pixels where simple rules don't seem to fit. CNNs are a good choice for Image Processing tasks as they are 2-Dimensional. Which removes the need of transforming Image data. We can train models to perform a single task or perform multiple tasks and produce an aggregated output as seen in One Shot 3D and Mask R-CNN. These Models are aiding across various domains such as Autonomous Driving, Computer Graphics, Medical Imagery etc. With these developments we can witness Computers Processing High level Visual information parallel to humans. Further advancements can help us bridge gaps in Human-Computer interaction, Robotics, and other similar domains.

REFERENCES

1. Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature, 2020; 585: 357–362. DOI: 10.1038/s41586-020-2649-2.
2. Van der Walt, S. et al. scikit-image: image processing in Python. PeerJ, 2014; 2: 453.
3. K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," IEEE International Conference on Computer Vision (ICCV), 2017; 2980-2988. doi:10.1109/ICCV.2017.322.
4. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39(6): 1137-1149. doi: 10.1109/TPAMI.2016.2577031.
5. Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell and Alexei A. Efros. Context Encoders: Feature Learning by Inpainting. In CVPR, 2016.
6. Kopf, J., Matzen, K., Alsisan, S., Quigley, O., Ge, F., Chong, Y., Cohen, M. One Shot 3D Photography. arXiv [cs.CV]. Opgehaal van http://arxiv.org/abs/2008.12298, 2020.
7. Kanopoulos, N., Vasanthavada, N., & Baker, R. L. Design of an image edge detection filter using the Sobel operator. IEEE Journal of Solid-State Circuits, 1988; 23(2): 358–367.
8. Canny, J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986; (6): 679–698.
9. S.B.Damelin and N.S.Hoang. “On Surface Completion and Image Inpainting by Biharmonic Functions: Numerical Aspects”, International Journal of Mathematics and Mathematical Sciences, 2018. Article ID 3950312 DOI:10.1155/2018/3950312.
10. Fig 1 Photo by Lily Banse on Unsplash.