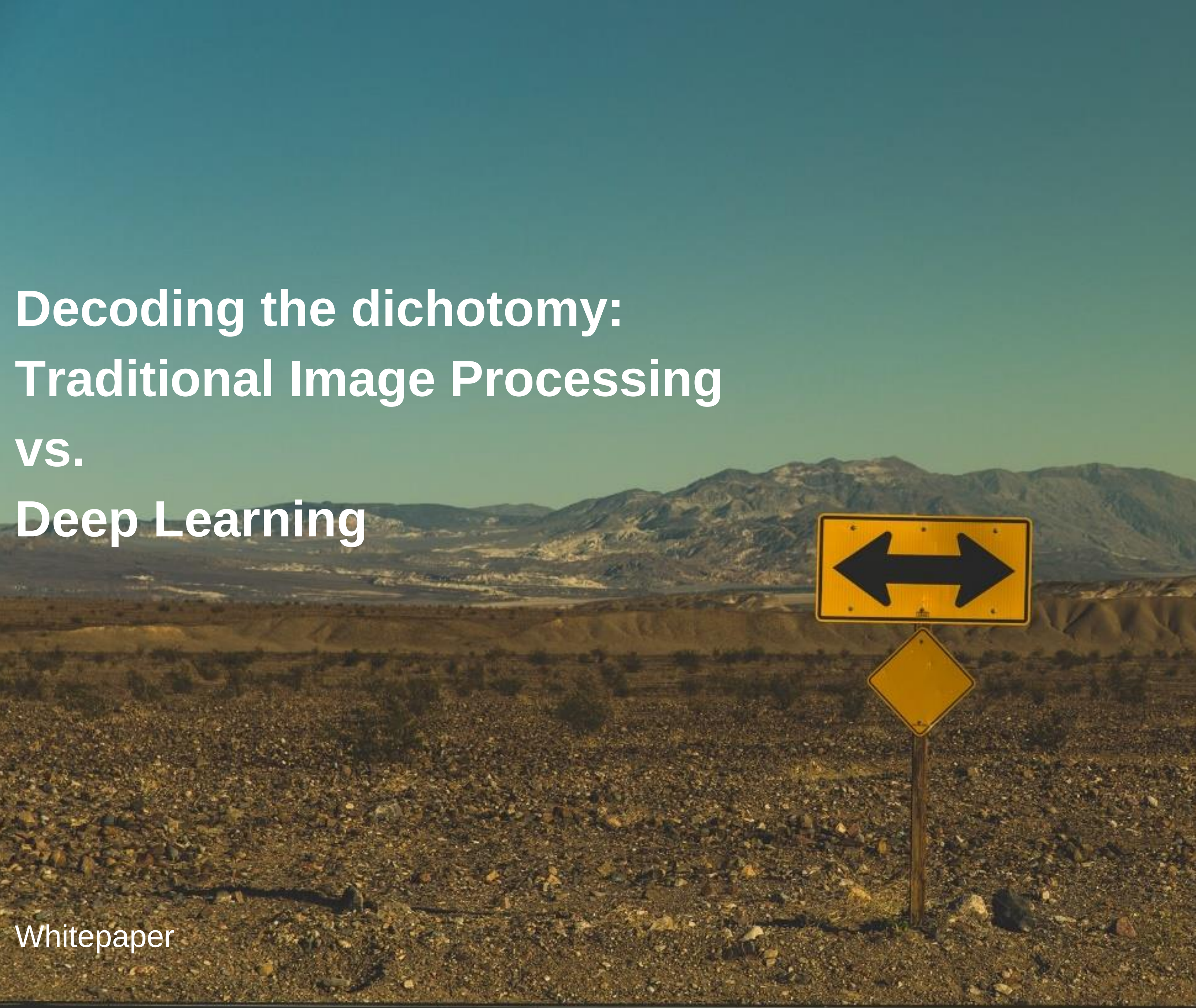
Decoding the dichotomy: Traditional Image Processing vs. Deep Learning



Whitepaper

Highlights

- Comparison of Deep Learning and Traditional Image Processing
- Hybrid Approaches
- Guidelines for making a suitable choice

Photo by Pablo García Saldaña on Unsplash

Note

For the context of this paper, the word “traditional image processing” shall be used to refer to a broader area of image processing which encompasses domains of image processing, computer vision, and classical machine learning.

1. Introduction

Deep learning has certainly revolutionized traditional image processing. It has pushed the boundaries of Artificial Intelligence to unlock potential opportunities across industry verticals.

In contrast to traditional image processing techniques, DL helps achieve greater accuracy in tasks such as object detection, image classification, Simultaneous Localization and Mapping (SLAM), and semantic segmentation.

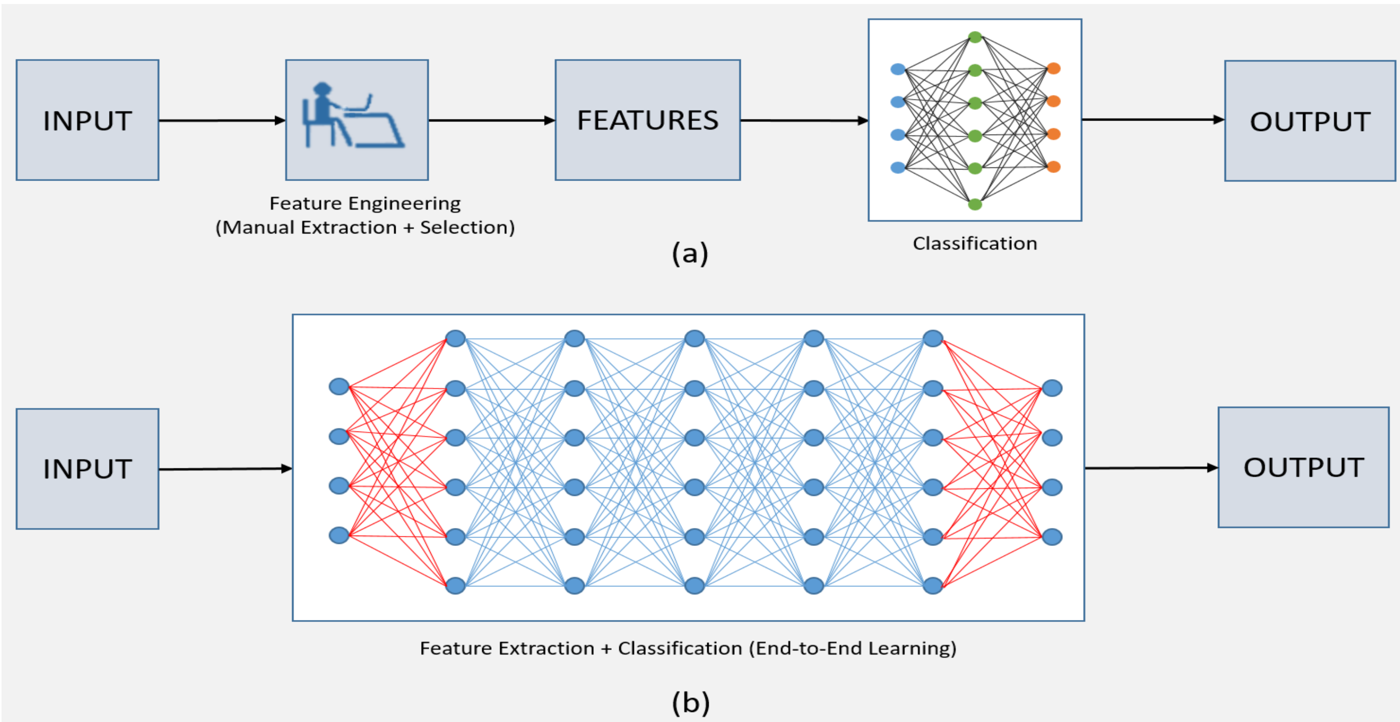


Figure 1: (a) Traditional Image Processing workflow vs. (b) Deep Learning workflow.

Several challenges that once seemed impossible to solve, are now solved to a point where machines are performing better than humans. However, that does not mean that the traditional image processing techniques that have advanced in the years before the rise of DL have been made obsolete.

This paper will analyze the benefits and drawbacks of each approach. This paper aims to provide better clarity on the subject which can help data scientists/ industries choose the most suitable method depending on the task at hand.

2. Comparison of Deep Learning and Traditional Image processing

2.1. Advantages of DL

Rapid advancement in DL and enhancements in device capabilities including memory capacity, computing power, power consumption, optics, and image sensor resolution have accelerated the spread of vision-based applications along with improved performance and cost-effectiveness.

Since neural networks used in DL are trained rather than programmed, applications following this approach often require less fine-tuning and expert analysis. The availability of a humongous amount of video data in today's system supports this cause. While CV algorithms tend to be more domain-specific, DL, on the other hand, provides superior flexibility because CNN models and frameworks can be re-trained using a custom dataset for any use case.

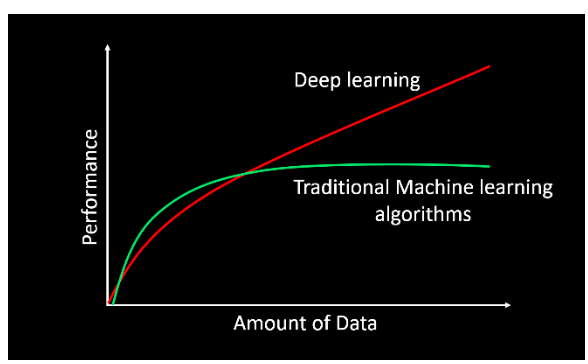


Figure 2: Data vs Performance Comparison

2.2. Advantages of traditional Image processing

At times, deep learning is overkill as traditional image processing can often solve a given problem with greater accuracy and in fewer lines of code than DL. The features learned from a deep neural net are specific to the training dataset which if not well constructed, probably won't perform well for images different from the training set. In contrast, algorithms like SIFT and even simple color thresholding and pixel counting algorithms are not class-specific, that is, they are very general and perform the same for any image.

Figure 3: Example of image stitching

Therefore, SIFT and other algorithms are often preferred for applications such as 3D mesh reconstruction/ image-stitching which do not need specific class knowledge. While the solutions to these tasks can be attained by training huge datasets, the vast research effort required for it is not feasible for a closed application. Summarily, deciding on the most suitable approach for a given computer vision problem, one should consider practical feasibility.

A product classification problem can be considered as an example. Supposing the problem aims to classify cans of food on a conveyor belt into either vegetarian or non-vegetarian distinguished by can color – green for veg. and red for non-veg. While the problem can be solved using accurate DL models generated by collecting sufficient training data, the traditional image processing is a much-preferred alternative in this scenario with its simple color thresholding technique. This example also highlights the fact that DL often fails to generalize the task at hand in the event of limited training dataset leading to over-fitting.

Manual tweaking of parameters of a model is a daunting task since a DNN consists of parameters in the order of millions inside it, each with complex interrelationships. As a result, DL models are censured to be a black-box. On the contrary, traditional image processing offers complete transparency and allows one a good estimate of how his/ her techniques will behave outside the training environment. It also offers flexibility to CV engineers to tweak their parameters to either improve their algorithm to achieve better accuracy and performance or investigate their mistakes when the algorithm fails. Traditional image processing is preferred for edge computing too, owing to its delivery of high performance with lower resource usage. This also makes traditional image processing more popular for cloud-based applications where high-powered resources that are required for deep learning applications are expensive.

2.3. Hybrid Approaches

Hybrid approaches are an amalgamation of traditional image processing and deep learning that present the best of both. They are gaining importance owing to their ability to maintain the right balance between mature and proven traditional image processing algorithms and versatile and accurate deep learning techniques.

Hybrid approaches have witnessed resounding success in medical image processing. Doctors can generally diagnose if a tumor is benign or malignant through mammal review, but hybridizing DL and CV capabilities allows us to automate this process and reduce the possibility of human error.

They are notably efficient in high-performance systems that require quick development. For example, an image processing algorithm can competently perform face detection over the live feed from a security camera. These detections can then be relayed to a DNN as the next stage for face recognition.

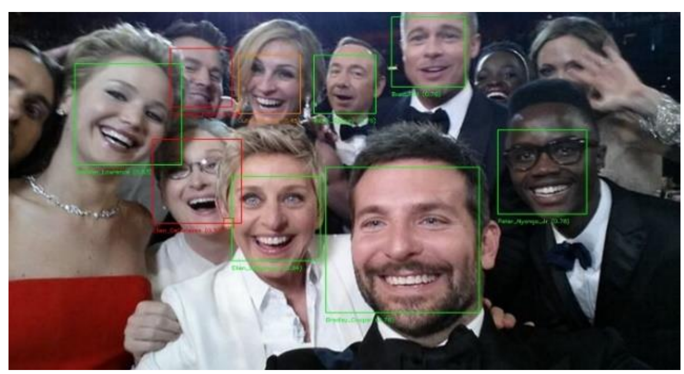


Figure 4: Example of hybrid approach – face recognition

This helps the DNN to work only on a small patch of the image thereby, reducing the considerable amount of computing resources and training effort that would otherwise have been required to process the entire frame.

Fusion can also help achieve better accuracy. One such classic example is document processing where traditional image processing techniques are used for pre-processing tasks like noise reduction, skew detection/ correction, and localization of lines and words. This, when followed OCR using deep techniques, yield better accuracy.

The blend of machine learning metrics and deep networks has gained significance over the years, owing to the evidence that it results in better models. Hybrid vision processing implementations have proved a performance advantage while providing a 130x-1,000x reduction in multiply-accumulate operations and about 10x improvement in frame rates compared to a pure DL solution. Furthermore, the hybrid implementation requires significantly lower CPU resources and about half of the memory bandwidth.

3. Guidelines for making a suitable choice

Andrew Ng, one of the famous AI practitioners once quoted: "The analogy to deep learning is that the rocket engine is the deep learning models and the fuel is the huge amounts of data we can feed to these algorithms." It can also be extrapolated to say that the quality of fuel (data) also plays a prominent role in the performance of the rocket (model accuracy).

Deep learning is certainly impressive and exciting, but it is not suitable for every situation. There are certain circumstances where deep learning is probably not the best solution. Popular conventions regard deep learning unsuitable for below use-cases:

- Augmented/ virtual reality
- 3D modeling
- Video stabilization
- Motion capture/ calculation
- Noise reduction
- Image registration
- Stereo processing
- Data compression and coding

DL triumphs at solving closed-end classification problems, which aims at mapping a range of potential signals to a finite number of categories, given that there is sufficient training data available and the test set does not deviate too much from the training set. However, an aberration from these assumptions leads to critical problems that DL often fails to acknowledge.

Popular object detection datasets provide stats about the data size that is required by deep learning for achieving good accuracy.

ImageNet – 15 million images with 1000 object categories/ classes
Open Images v6 – 9 million images, with 600 object categories/ classes
Microsoft Common Objects in Context (COCO) – 2.5 million images, 91 object categories
PASCAL VOC dataset – 500K images, 20 object categories

Another point to be noted is the difficulty faced by ML models in dealing with priors. It simply refers to the fact that not everything can be learned from data which requires more priors to be injected into the models. Applications associated with 3D CV such as image-based 3D modeling depict such a scenario. For optimal performance, such problems require strong priors such as smoothness, silhouette, and illumination information. CV approaches can either work independently or be complemented and work with DL for the following applications: morphing/ blending, sharpening, optical corrections/ transformations, calculating geometries, segmentation, de-blurring.

The below guidelines summarizes the common attributes of each technology from the preceding discussions. These guidelines also act as a handy tool for not only data scientists but novice developers and business people who do not necessarily have a thorough understanding of the subject to make better decisions.

Prefer Deep Learning when:

- Huge training data available for making accurate decisions.
- Possess high-computing power (i.e., CPU, GPU, TPU, etc.) - to allow intensive model training and good application performance.
- Uncertain about the positive feature-engineering outcome (i.e., selecting the most suitable feature(s) yielding the desired outcome), especially in unstructured media (audio, text, images).
- Deployment restricted to high-performance devices (i.e., unsuitable for embedded, micro-controllers).
- Less/ no domain expertise is available

Stick to traditional image processing when:

- Limited (annotated/ labeled) data available.
- Lack of high storage and computing power.
- Cheaper solution desired
- Desire flexible deployment over a range of hardware.
- Good domain expertise present.

The below table summarizes the comparison between deep learning and traditional image processing:

Selection criteria	Traditional Image Processing	Deep Learning
Training dataset	Small	Large
Computing power	Low	High
Feature engineering	Required	Unnecessary
Training time	Short	Long
Annotation time	Short	Long
Algorithm Transparency	High	Low
Domain expertise	High	Low
Priors (Assumptions)	Few	Many
Proprietary material - Risk of exposure	DLLs (Risk – Negligible)	Model files, DLLs (Risk – High)
Deployment flexibility	High	Low
Expenditure (BOM)	Low	High

Some typical applications of DL and traditional image processing are given below:

Traditional Image Processing	Deep Learning
Image transformation (Lens distortion correction, view changes)	Image classification (OCR and Handwritten character recognition)
Image Signal Processing (ISP)	Object detection/ identification
Camera calibration	Semantic segmentation
Industrial inspection – Defect detection	Instance segmentation
Stereo image processing	Image synthesis
Automatic panorama stitching	Image colorization
3D data processing	Image Super-resolution
Calculating geometries	Scene understanding

4. Conclusion

This paper presented arguments for why traditional image processing techniques are still very much relevant in this age of Deep Learning.

Deep learning can be seen as one of the tools to solve image/ video processing problems along with other programming languages, image processing techniques, classifiers, and machine learning. Some typical applications were also compared from the perspectives of traditional image processing and DL and discussed how the former is often a simpler alternative in tasks where DL is an overkill.

The paper also highlighted some areas which hint that hybrid approaches are the way forward with their lower memory bandwidth and computation requirements, and high performance and accuracy. We also witnessed some examples where traditional image processing techniques improve DL performance (i.e., reduced training time and lower processing power).

Some general guidelines were also discussed in the paper which shall prove to be useful for beginners as well as experts in the computer vision domain to make better decisions while choosing between deep learning and traditional image processing to solve their problem.

5. References

[1] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in Science and Information Conference. Springer,2019, pp. 128–144

[2] Zdziarski Z. (2018, January 26), The Reasons Behind the Growth of Computer Vision. *Zbigatron*. Retrieved from https://zbigatron.com/the-reasons-behind-the-recent-growth-of-computer-vision/

[3] Shaikh F. (April 8, 2017), Deep Learning vs. Machine Learning – the essential differences you need to know. *Analytics Vidhya*. Retrieved from https://www.analyticsvidhya.com/blog/2017/04/comparison-between-deep-learning-machine-learning/

[4] AI vs Machine Learning vs Deep Learning: What's the difference?. *Guru99.* Retrieved from https://www.guru99.com/machine-learning-vs-deep-learning.html

[5] Brownlee J. (2019, March 13). 9 Applications of Deep Learning for Computer Vision. *Machinelearningmastery*. Retrieved from https://machinelearningmastery.com/applications-of-deep-learning-for-computer-vision/

[6] Tulluri R. (2017, November 29). Conventional computer vision coupled with deep learning makes AI better. *Network World*. Retrieved from https://www.networkworld.com/article/3239146/conventional-computer-vision-coupled-with-deep-learning-makes-ai-better.html