FROM THE EYES OF A MACHINE: IMAGE RECOGNITION TECHNOLOGIES

Susan Pratt*

CITE AS: 5 GEO. L. TECH. REV. ___ (2021)

TABLE OF CONTENTS

I. INTRODUCTION

Image recognition technologies (IRTs) are tools that are capable of taking images, detecting and classifying objects within those images, and outputting some desired result based on the analysis. With computers becoming more powerful and image data availability increasing, IRTs are continually improving, enabling humans to use them for applications like optical character recognition (OCR), healthcare, and facial recognition.[1] IRTs

* Georgetown University Law Center, J.D. Candidate 2022; Yale University, Ph.D. in Physics 2019; University of Rochester, B.S. in Physics 2012 & B.A. in Mathematics 2012.

[1] *See, e.g.*, Naveen Joshi, *The Present and Future of Computer Vision*, FORBES (June 26, 2019, 11:19 PM), https://www.forbes.com/sites/cognitiveworld/2019/06/26/the-present-and-future-of-computer-vision/#26cba0f6517d [https://perma.cc/QT7X-DZWF]; Kaz Jaszczak, *Optical Character Recognition: A Backbone for Postal and Mail Sorting Applications*, MAILING SYSTEMS TECH. (Apr. 7, 2011, 10:18 AM), https://mailingsystemstechnology.com/article-

are numerous and complex, and they have a growing presence in the image recognition market.[2] A brief overview of IRTs is provided, followed by an introduction to the machine learning tools used to implement IRTs, and an overview of their advantages and disadvantages.

II. DEFINING IRTS

A computer "sees," detects, and recognizes objects in an image, and makes decisions on what to do with the information obtained from the analyzed image.[3] To a computer, an image is a grid or, more technically, a matrix of values where each value represents a single pixel of the image. The pixels in a grayscale image, or an image depicting only shades of gray, can take values from 0 to 255.[4] For a colored image, the matrix is three-dimensional, where each dimension corresponds to a single-color channel—red, green, or blue—and the pixels in each channel can take values from 0 to 255.[5] Analyzing colored images is more complex than analyzing grayscale images because of the extra dimensionality of the data.[6]

When a computer is given an image to analyze, it may discern certain features in the image, which can be used to identify and localize the objects of interest.[7] Suppose there is an image of three filled-in, non-touching, black circles against a white background. The goal might be to count the number of black circles in this image. Each pixel in the image will correspond to either the white background, which would have pixel values of 255, or the black circles with pixel values of 0. To determine the number of circles present, one then needs to find the number of closed, zero-pixel-value regions. The computer would be instructed to scan along each row and column of the image

---

2813-Optical-Character-Recognition-A-Backbone-for-Postal-and-Mail-Sorting-Applications.html [https://perma.cc/DQL7-CLNA].

[2] *See, e.g.*, *Image Recognition Applications: 7 Essential Future Uses*, DATA SCIENCE SOCIETY, https://www.datasciencesociety.net/image-recognition-applications-7-essential-future-uses/ [https://perma.cc/3SAD-G7BB]. *See generally* Joshi, *supra* note 1.

[3] *See, e.g.*, CHRISTOPHER M. BISHOP, PATTERN RECOGNITION AND MACHINE LEARNING 1–2 (2006).

[4] Aymane Hachcham, *Exploring Image Processing Techniques – OpenCV*, TOWARDS DATA SCIENCE (Apr. 29, 2020), https://towardsdatascience.com/exploring-image-processing-techniques-opencv-4860006a243 [https://perma.cc/X5SW-GM7W].

[5] *Id.*

[6] *Id.*

[7] *See* Kevin Murphy, Antonio Torralba & William T. Freeman, *Using the Forest to See the Trees: A Graphical Model Relating Features, Objects, and Scenes*, 16 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 1499, 1499–1500 (S. Thrun, L. Saul, & B. Scholköpf eds., 2003), https://proceedings.neurips.cc/paper/2003/file/99f59c0842e83c808dd1813b48a37c6a-Paper.pdf [https://perma.cc/ZTK6-YJ7H].

matrix and to take note of whenever the pixel value changes from 0 to 255, or vice versa. The changes in pixel values will correspond to the edges of the individual circles, allowing one to create a simple program to use this knowledge of edge features to determine the number of black circles present. In a different approach, the protocol might be to find the minima of the pixel values here and locate the centers of these minima. The example of the black circles against a white background is a simple one, but it builds a foundation from which IRTs are derived. To summarize, images are made of pixels, the computer "sees" those pixels, and features like points or edges can be extracted from the images to help identify and localize objects of interest.

IRTs, however, are interested in much more complex images than noiseless, black circles against white backgrounds.[8] Analyzing complex images for purposes like skin cancer detection requires more advanced approaches in the realm of machine learning.[9]

A. Machine Learning

Machine learning refers to the ability of computers to learn and improve what they output based on what they learned. More specifically, machine learning entails using algorithms to train computers to identify patterns such that when new data is provided post-training, the machine can identify or predict this data given what it learned earlier.[10] Machine learning is broken down into several categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.[11]

1. Supervised Learning

Suppose there is a model that can predict whether it is day or night based on how much light there is outside. In supervised learning, the data that is used to train the model is called "labeled data," because given a certain input

---

[8] *See Image Recognition Applications: 7 Essential Future Uses*, *supra* note 2.

[9] *See* Kevin Vu, *Beginner's Guide: Image Recognition and Deep Learning*, DZONE (Nov. 29, 2018), https://dzone.com/articles/beginners-guide-image-recognition-and-deep-learnin [https://perma.cc/JCN7-FC27].

[10] *See* Karen Hao, *What is Machine Learning?*, MIT TECH. R. (Nov. 17, 2018), https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/ [https://perma.cc/GDT9-VRUC]; CHRISTOPHER M. BISHOP, PATTERN RECOGNITION AND MACHINE LEARNING 1–2 (2006); Javaid Nabi, *Machine Learning – Fundamentals*, TOWARDS DATA SCIENCE (Aug. 15, 2018), https://towardsdatascience.com/machine-learning-basics-part-1-a36d38c7916 [https://perma.cc/DGU5-W6SP].

[11] CHRISTOPHER M. BISHOP, PATTERN RECOGNITION AND MACHINE LEARNING 3 (2006).

value the user knows the output value.[12] For the day/night problem, the trained data would be light measurements for which it is already known whether the light quantities fall into the category of "day" or "night." This labeled data is used to train the computer model until the model reaches a certain level of accuracy. Once the model is complete, a user can then provide new light measurement data and have the computer determine whether the values fall into the category of "day" or "night."

2. Unsupervised Learning

In unsupervised learning, the data is not labeled,[13] meaning it is unknown what quantities of light qualifies as "day" or "night." The models are developed by taking the training data and searching for structures within the data.[14] For the light example, a computer might determine that the light values provided to it can be separated into two clusters of values. When provided with a raw light measurement, the computer could then determine if it is more likely to be in one cluster over another.

3. Semi-supervised Learning

Semi-supervised learning is a combination of supervised and unsupervised learning.[15] The training data is a mixture of labeled and unlabeled data and the machine is initially partially trained using the labeled data.[16] The partially-trained model can then be used to label the unlabeled data to obtain pseudo-labeled data, and the new set of labeled and pseudo-labeled data are used to train the model.[17]

4. Reinforcement Learning

Reinforcement learning is an approach where the computer is given a starting point and a desired endpoint, and the machine is tasked with reaching

---

[12] *Id.*
[13] *Id.*
[14] *Id.*
[15] See F.D., *Simple Explanation of Semi-Supervised Learning and Pseudo Labeling*, TOWARDS DATA SCIENCE (Oct. 10, 2017), https://towardsdatascience.com/simple-explanation-of-semi-supervised-learning-and-pseudo-labeling-c2218e8c769b [https://perma.cc/M7QU-LQQ5].
[16] *Id.*
[17] *Id.*

the endpoint through a system of rewards and punishments.[18] The goal is to reach the endpoint or to solve a problem while achieving the maximum reward possible.[19]

B. Deep Learning

Machine Learning can be used to take a substantial number of images, which are labeled, unlabeled, or a mixture of labeled and unlabeled, and generate a model that can be used to make decisions, like classifying objects within an image or determining the probability that objects in an image belong to certain classes. Deep learning is a subcategory of machine learning and forms the basis of many of the IRTs available or in development.[20] It is grounded in principles learned through neuroscience regarding how information is processed in mammalian brains.[21] Deep learning does not require manual feature extraction; instead, it uses neural networks to ask questions about the images in the training data set and subsequently classifies the data.[22]

One such neural network is called a convolutional neural network (CNN).[23] CNNs use a layered approach to analyze images. The layers generally include a convolution layer, a Rectified Linear Unit layer (ReLU), a pooling layer, and a fully connected layer.[24] In a convolution layer, a filter that is much smaller in size than the input image and consisting of values correlating to some feature like a line is convolved with the input image; that

---

[18] *See, e.g.*, Błażej Osiński & Konrad Budek, *What is Reinforcement Learning? The Complete Guide*, DEEPSENSE.AI (July 5, 2018) https://deepsense.ai/what-is-reinforcement-learning-the-complete-guide/ [https://perma.cc/E2HG-JCF7]; CHRISTOPHER M. BISHOP, PATTERN RECOGNITION AND MACHINE LEARNING 3 (2006).

[19] *Id*.

[20] Vu, *supra* note 9.

[21] *See, e.g.*, Itamar Arel, Derek C. Rose & Thomas P. Karnowski, *Deep Machine Learning – A New Frontier in Artificial Intelligence Research [Research Frontier]*, 5 IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE 13, 13 (2010) http://web.eecs.utk.edu/~ielhanan/Papers/CIM2010.pdf [https://perma.cc/B8NJ-5F89].

[22] Shivam Arora, *5 Major Differences Between Machine Learning and Deep Learning*, SIMPLILEARN (Sept. 25, 2020) https://www.simplilearn.com/machine-learning-vs-deep-learning-major-differences-you-need-to-know-article [https://perma.cc/KZ2R-XGEU].

[23] *See generally* Luis Serrano, *A Friendly Introduction to Convolutional Neural Networks and Image Recognition*, YOUTUBE (Mar. 20, 2017), https://www.youtube.com/watch?v=2-Ol7ZB0MmU [https://perma.cc/SST5-77RG]; Yann LeCun et al., *Backpropagation Applied to Handwritten Zip Code Recognition*, 1 Neural Computation 541 (1989) http://yann.lecun.com/exdb/publis/pdf/lecun-89e.pdf [https://perma.cc/XFF8-K6TR].

[24] *See, e.g.*, Afshine Amidi & Shervine Amidi, *Convolutional Neural Networks Cheatsheet*, CS 230 – DEEP LEARNING, https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks [https://perma.cc/ED2J-TDQQ].

is, it undergoes a multiplication-like operation with the input image.[25] The filter is systematically moved and convolved with the input image until the entire image has undergone this process.[26] This layer helps detect features.[27] The output from the convolution layer will then have an activation function applied to it in the ReLU layer, where positive values are kept as they are and all other values are set to zero.[28] The activation function fundamentally determines what information is relevant and impacts the output of the neural network.[29] The pooling layer downsamples the output of the convolution and ReLU layers, compressing the information.[30] Similar to the convolution process with a filter, the pooling process uses a small filter-like window, sliding it systematically around the output of the convolution and ReLU layers.[31] The window here might be used to find the maximum value ("max pooling") in the region with which it overlaps, and the maximum value will be mapped to the output.[32] The fully connected layer takes the output from the max pooling layer and flattens the information into a single vector of values where each represents the probability a feature corresponds to a particular label.[33] The CNN can be optimized by altering aspects of the model, like the number of filters used in the convolution layers, the number of layers included in the system, or the designation of which layers are repeated.[34]

Deep learning neural networks are considered "deep" because they possess many layers in a model. The deepness allows for the ability to extract low-level (e.g., lines and circles), mid-level (e.g., lines and circles forming eyes and noses), and high-level (e.g., lines and circles forming eyes and noses

---

[25] *Id.*
[26] *Id.*
[27] *Id.*
[28] *Id.*
[29] Dinesh Kumawat, *7 Types of Activation Functions in Neural Network*, ANALYTICS STEPS (Aug. 22, 2019), https://www.analyticssteps.com/blogs/7-types-activation-functions-neural-network [https://perma.cc/48Q5-RRHD]; *CS231n Convolutional Neural Networks for Visual Recognition*, https://cs231n.github.io/neural-networks-1/ [https://perma.cc/7GRZ-WLWF].
[30] Amidi & Amidi, *supra* note 24.
[31] *Id.*
[32] *Id.*
[33] *Fully Connected Layers in Convolutional Neural Networks: The Complete Guide*, missinglink.ai, https://missinglink.ai/guides/convolutional-neural-networks/fully-connected-layers-convolutional-neural-networks-complete-guide/ [https://perma.cc/CK5D-C9ZW].
[34] *See* Jason Brownlee, *Convolutional Neural Network Model Innovations for Image Classification*, MACHINE LEARNING MASTERY (April 24, 2019) https://machinelearningmastery.com/review-of-architectural-innovations-for-convolutional-neural-networks-for-image-classification/ [https://perma.cc/42YW-WJVY] (last updated July 5, 2019).

forming faces) features from images.[35] Some examples of deep neural network models are VGG-16 with 16 layers, GoogLeNet with 22 layers, and ResNet with at least 34 layers.[36]

III. ADVANTAGES AND DISADVANTAGES OF IRTS

A. Advantages

Advantages of IRTs include saving resources like time and money, having greater accuracy than humans, and improving quality of life.[37] Using machine learning algorithms, IRTs can handle large quantities of data.[38] Imagine being given millions of images and tasked with categorizing the data. Such a task requires an immense amount of human effort and would be costly. IRTs save time and resources on such tasks.[39] One type of IRT is optical character recognition (OCR). This technology is used to identify text from images, and it is indispensable in the sorting of postal mail.[40] OCR technology is capable of taking an image of an address that is handwritten or typed and converting that image of text to digitized text which can be used to automatically sort mail.[41] OCR is also utilized in areas like automated license

---

[35] Luis Serrano, A Friendly Introduction to Convolutional Neural Networks and Image Recognition, YouTube (Mar. 20, 2017), https://www.youtube.com/watch?v=2-Ol7ZB0MmU [https://perma.cc/SST5-77RG].

[36] Vivienne Sze et al., *Efficient Processing of Deep Neural Networks: A Tutorial and Survey*, 105 PROCEEDINGS OF THE IEEE 2295, 2304–05 https://ieeexplore.ieee.org/document/8114708 (last visited Nov. 8, 2020).

[37] *See, e.g.*, *How Facial Recognition Technology will Save Companies Time and Money*, TEEM (May 18, 2017), https://www.teem.com/blog/facial-recognition-technology [https://perma.cc/JWD2-8H3E] (discussing how facial recognition technologies may save time and money); Andre Esteva et al., *Dermatologist-level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115, 115, 117–18 (2017), https://www.nature.com/articles/nature21056 (last visited Nov. 8, 2020) (showing a CNN outperforming human dermatologists); Amit Chowdhry, *How Facebook is Helping Blind People 'See' Photos*, FORBES (April 7, 2016, 11:44 PM) https://www.forbes.com/sites/amitchowdhry/2016/04/07/facebook-automatic-alternative-text/?sh=6ed92ced7c86 [https://perma.cc/G97D-RSNT] (developing a program to help the blind "see" Facebook pictures).

[38] *See* Matthew Stewart, *Handling Big Datasets for Machine Learning*, TOWARDS DATA SCIENCE (Mar. 11, 2019) https://towardsdatascience.com/machine-learning-with-big-data-86bcb39f2f0b [https://perma.cc/FPZ2-V3NE].

[39] *See id.*

[40] Kaz Jaszczak, *Optical Character Recognition: A Backbone for Postal and Mail Sorting Applications,* MAILING SYSTEMS TECH (April 7, 2011, 10:18 AM), https://mailingsystemstechnology.com/article-2813-Optical-Character-Recognition-A-Backbone-for-Postal-and-Mail-Sorting-Applications.html [https://perma.cc/DQL7-CLNA].

[41] *Id.*

plate readers (ALPRs), where law enforcement can capture images of automotive license plates. OCR technology converts the license plate images into alphanumeric characters that can subsequently be compared to license plate numbers in a database.[42] ALPRs reduce the amount of time it would take an officer to obtain a license plate number, enter the alphanumeric characters in a database, and determine if the plate belongs to a vehicle of interest.[43]

IRTs are becoming more accurate in image detection in comparison to humans.[44] This is relevant for medical detection, where early detection is crucial for early diagnoses of diseases like skin cancer.[45] There remains much work and research to bring IRTs to the level where human supervision is not needed to fully operate along with the IRT for medical purposes. However, some IRTs already exist on the market for medical detection. One example is cmTriage™, which uses machine learning to aid radiologists in prioritizing cases that may present suspicious findings related to breast cancer.[46] cmTriage™ received FDA clearance to operate as a "notification-only, parallel workflow test."[47] Another example is a mobile phone application called SkinVision, which uses a CNN to provide the user with a risk assessment of an imaged skin patch.[48] The application labels the imaged spot by skin cancer risk level.[49]

IRTs can improve human experiences. In 2016, Facebook launched Automatic Alt Text, which enabled visually impaired or blind users to "see"

---

[42] *See, e.g.*, Pam Greenberg, *Automated License Plate Readers* (Feb. 2015), NAT'L CONF. OF STATE LEGIS., https://www.ncsl.org/research/telecommunications-and-information-technology/automated-license-plate-readers.aspx [https://perma.cc/MDH2-QS6Z]; DAVID J. ROBERTS & MEGHANN CASANOVA, AUTOMATED LICENSE PLATE RECOGNITION SYSTEMS: POLICY AND OPERATIONAL GUIDANCE FOR LAW ENFORCEMENT 1 (2012), https://www.ncjrs.gov/pdffiles1/nij/grants/239604.pdf [https://perma.cc/8R2C-TTEA]; *Automated License Plate Recognition*, IACP, https://www.theiacp.org/projects/automated-license-plate-recognition [https://perma.cc/8CGD-STJ8].

[43] *See, e.g.*, DAVID J. ROBERTS & MEGHANN CASANOVA, AUTOMATED LICENSE PLATE RECOGNITION SYSTEMS: POLICY AND OPERATIONAL GUIDANCE FOR LAW ENFORCEMENT 1 (2012).

[44] Esteva et al., *supra* note 37.

[45] *Id.* at 115.

[46] *cmTriage™*, CUREMETRIX, https://curemetrix.com/cm-triage-2/ [https://perma.cc/QP9U-6G9D].

[47] *Id.*

[48] *What is SkinVision?*, SKINVISION, https://skinvision.zendesk.com/hc/en-us/articles/207999609-What-is-SkinVision- [https://perma.cc/3E27-379P]; *How Does the Risk Assessment Work?*, SKINVISION, https://skinvision.zendesk.com/hc/en-us/articles/115003114509-How-does-the-risk-assessment-work- [https://perma.cc/KU83-WUG8].

[49] *How Does the Risk Assessment Work?*, SKINVISION, https://skinvision.zendesk.com/hc/en-us/articles/115003114509-How-does-the-risk-assessment-work- [https://perma.cc/KU83-WUG8].

Facebook pictures by converting images to a text description that could be heard, using a text-to-speech tool.[50] In 2017, Apple introduced Face ID, a facial recognition technology using neural networks, allowing mobile phone users to use their faces as a means for unlocking their phones and more.[51]

B. Disadvantages

IRTs are error-prone, may present privacy concerns, and may be used to target certain communities for surveillance purposes.[52] Computers using machine learning-based algorithms are still not capable of perceiving elements in real-world images in the same way that humans are capable of doing and may make mistakes.[53] For example, a study showed that a "stop" sign could be seen and determined by a deep neural network to be a "speed limit 45" sign.[54]

Additionally, although IRTs are being used for the benefit of society, there are concerns that some IRT applications may lead to privacy violations and concerns that minority populations may be targeted.[55] These concerns relate primarily to facial recognition technology. For example, alarms were raised when a statement was made that a researched facial recognition tool could "predict criminals."[56] In addition to facial recognition, automated license plate readers pose similar privacy and bias concerns.[57] In 2012, the

[50] Chowdhry, *supra* note 37.
[51] *About Face ID Advanced Technology*, APPLE INC., https://support.apple.com/en-us/HT208108 [https://perma.cc/UX77-TPGN]; *How Does Apple's New Face ID Technology Work?*, FORBES (Sept. 13, 2017, 12:47 PM), https://www.forbes.com/sites/quora/2017/09/13/how-does-apples-new-face-id-technology-work/?sh=502345a42b7f [https://perma.cc/5TP8-9C6T].
[52] *See* Douglas Heaven, *Why Deep-Learning AIs are so Easy to Fool*, NATURE (Oct. 9, 2019), https://www.nature.com/articles/d41586-019-03013-5 [https://perma.cc/H4N3-2TWD] (discussing "the potential for sabotaging AI"); Ángel Díaz & Rachel Levinson-Waldman, *Automatic License Plate Readers: Legal Status and Policy Recommendations for Law Enforcement Use*, BRENNAN CENTER FOR JUSTICE (Sept. 10, 2020) https://www.brennancenter.org/our-work/research-reports/automatic-license-plate-readers-legal-status-and-policy-recommendations [https://perma.cc/9DUW-6BTX] (discussing the "high error rates" of ALPRs, as well as the privacy and disparate impact concerns of this technology).
[53] *See generally* Heaven, *supra* note 52.
[54] Kevin Eykholt et al., Robust Physical-World Attacks on Deep Learning Visual Classification, CVPR 1625, 1630 (2018), https://openaccess.thecvf.com/content_cvpr_2018/papers/Eykholt_Robust_Physical-World_Attacks_CVPR_2018_paper.pdf [https://perma.cc/7U9C-B5NF].
[55] *See, e.g.*, Díaz & Levinson-Waldman, *supra* note 52.
[56] *Facial Recognition to 'Predict Criminals' Sparks Row Over AI Bias*, BBC NEWS, (June 24, 2020), https://www.bbc.com/news/technology-53165286 [https://perma.cc/J8QS-TER7].
[57] *See, e.g.*, Díaz & Levinson-Waldman, *supra* note 52.

New York Police Department admitted to collecting targeted ALPR data (and other information) on worshippers of Muslim mosques, which led to an outcry that such targeting was occurring.[58]

IV. CONCLUSION

As science and technology advances, IRTs are becoming more advanced and capable of handling increasingly complex image-related tasks. To understand what an IRT is, one must first understand how computers "see." Once it was better understood how computers "see" and how visual information is processed by mammals, powerful machine learning algorithms were developed to create technologies that could be provided with images and tasked with detecting or classifying elements within the images. Decisions could then be based on the outputs of the algorithms. IRTs like ALPRs, skin cancer detection mobile phone apps, and facial recognition technologies evolved. Although IRTs may pose privacy and bias concerns, the current and possible future benefits they bring to areas like healthcare illustrate their enormous potential to have a positive impact on society.

[58] Adam Goldman & Matt Apuzzo, *NYPD Defends Tactics Over Mosque Spying; Records Reveal New Details on Muslim Surveillance*, HUFFPOST, https://www.huffpost.com/entry/nypd-defends-tactics-over_n_1298997 [https://perma.cc/LG7U-55LW].