VISUAL PERCEPTION OF THE MACHINE

information design of
Convolutional Neural Network

Yueling Wu

Candidate: Yueling Wu
Matricola: 903488

Thesis Supervisor: Michele Mauri
Co-Supervisor: Beatrice Gobbo

Master degree in Communication Design
Academic Year 2020/2021
15 November 2020

POLITECNICO
MILANO 1863

ABSTRACT

ENGLISH VERSION

Convolutional Neural Network is a class of powerful algorithms of data processing, notably used in the field of image recognition. The underlying processing models are often opaque like black boxes and difficult to be understood by people who are not professions in those specific fields. The research aims both to visually explain the algorithm itself, and to discuss the factors that led to algorithmic biases. These two goals are reached through communication design methods. By presenting data visualisations and diagrams, it allows the general audience to rely on their collective experience to comprehend and criticise this kind of algorithm. This research starts reviewing Computer Science knowledge, analysing the most relevant publications and existing visual interpretations created by experts. Based on the analysis of the target users, a visual explanation has been designed combined with applications related to daily lives and algorithmic errors.

ABSTRACT

VERSIONE ITALIANA

Convolutional Neural Network è una classe di potenti algoritmi di elaborazione dei dati, utilizzati in particolare nel campo del riconoscimento delle immagini. I modelli di elaborazione sottostanti sono spesso opachi come scatole nere e difficili da comprendere da persone che non sono professioni in quei campi specifici. La ricerca mira sia a spiegare visivamente l'algoritmo stesso, sia a discutere i fattori che hanno portato a bias algoritmici. Questi due obiettivi vengono raggiunti attraverso metodi di design della comunicazione. Presentando visualizzazioni di dati e diagrammi, consente al pubblico generale di fare affidamento sulla propria esperienza collettiva per comprendere e criticare questo tipo di algoritmo. Questa ricerca inizia esaminando la conoscenza dell'informatica, analizzando le pubblicazioni più rilevanti e le interpretazioni visive esistenti create da esperti. Sulla base dell'analisi degli utenti target, è stata progettata una spiegazione visiva combinata con applicazioni relative alla vita quotidiana e agli errori algoritmici.

PREFACE

We are now in an era where programming increasingly relies on Machine Learning techniques [1]. Concepts like Machine Learning and Deep Learning are getting closer to the average user. Media headlines mention them, new products from Internet companies mention them. From shopping recommendations, personalised content recommendations, targeted advertising, loan assessments, insurance assessments, employee assessments, and criminal risk assessments in the judicial process, algorithms are making more and more decisions.

However, the problems also surfaced: artificial intelligence in tech companies discriminates against women in résumé screening [2], face recognition errors are far more common for people of darker colours than for people of lighter [3]. Algorithmic bias is a problem that needs to be faced. The issues of non-transparency, inaccuracy, unfairness and difficulty to review brought about by the code of rules need to be considered and studied carefully.

This research, taking the ConvNet (a common abbreviation for ‘Convolutional Neural Network’) recognition image as an example, aims to introduce the neural network to the lay user and explore some of the factors that lead to algorithmic bias in image recognition.

[1] The A-Z of AI: Machine Learning

[2] Amazon scraps secret AI recruiting tool that showed bias against women

[3] How We Made AI As Racist and Sexist As Humans

BACKGROUND

1.1 What is Convolutional Neural Network?

You may have been exposed to programming and developed one or two programs. At the same time, you have probably read overwhelming reports about Deep Learning or Machine Learning, although many times they are given a broader name: Artificial Intelligence. In layman's terms, Machine Learning is a discipline that discusses a variety of functional forms that can be applied for different problems and how data can be used effectively to obtain specific values of function parameters.

Deep Learning refers to a class of functions in Machine Learning, usually in the form of multi-layer neural networks. In recent years, relying on large data sets and powerful hardware, Deep Learning has become the primary method for processing complex, high-dimensional data, such as images, text Corpora, and sound signals.

In Deep Learning, ConvNet is a class of deep neural networks, most commonly applied to analysing visual imagery. (Valueva et al., 2020)

A convolutional network is a neural network specifically designed to process data with known, grid-like topology. For example, time-series data is a type of one-dimensional grid sampled at a specific time interval; image data is one of the kinds of two-dimensional pixel grid: each picture is a pixel matrix. ConvNets have achieved great success in practical applications, such as Siri, Face ID, face-modification apps, OCR, etc.

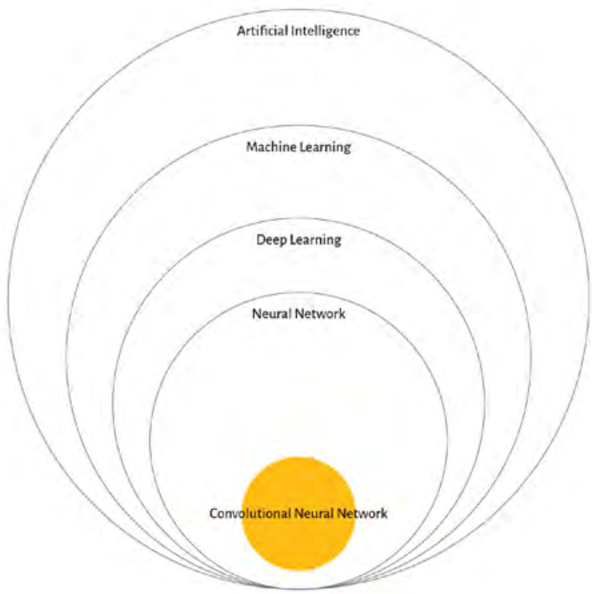


Figure. 1 The location of the Convolutional Neural Network in the AI knowledge system

1.2 Summarise the visual interpretations of Deep Learning Techniques

As Deep Learning becomes widespread in all areas, it is critical to empower users with tools to understand when models work correctly, when they fail, and ultimately how to improve performance, so that explainable Deep Learning is gaining attention. (Hohman et al., 2018) Computer scientists are also working on varieties of toolkits and visual explanations for better control Machine Learning techniques. Through Figure. 2, we can find that the explanations of Deep Learning are mainly carrying out among professions, especially from the questions “WHAT” and “HOW”: experts use scientific methods in their contexts to disseminate knowledge and share experiences.

Lay user lack expertise to understand these explanations, yet in real-world scenarios, they often encounter algorithmic errors. Nevertheless, because of the opacity of the Algorithm, users who know only input and output can not evaluate the algorithm's performance then suggest improvements.

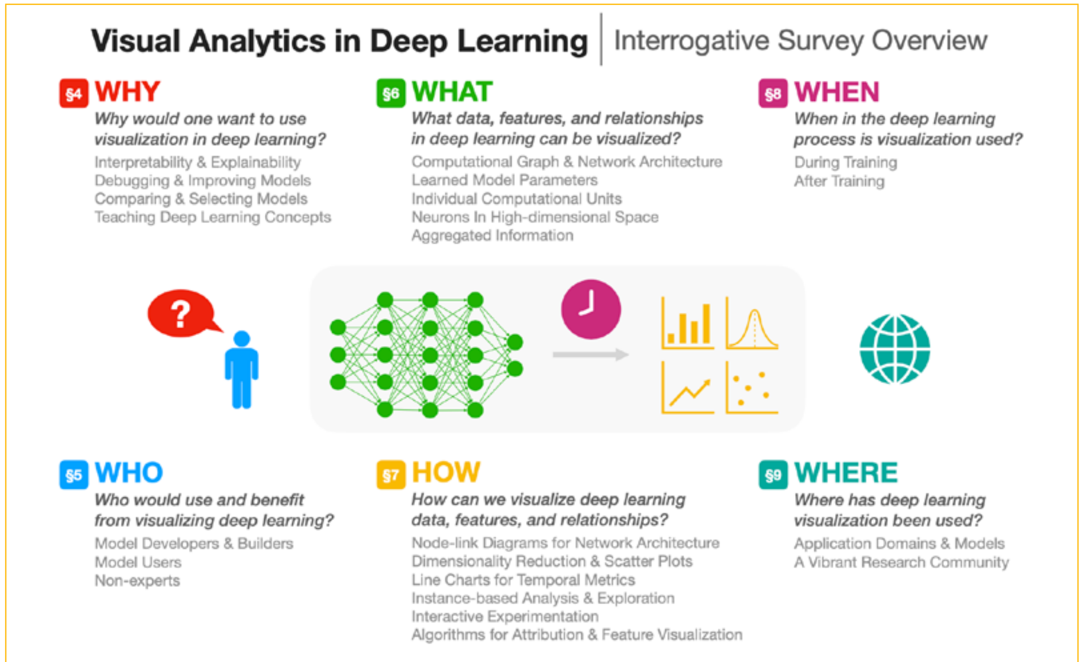


Fig. 1. A visual overview of our interrogative survey, and how each of the six questions, "Why, Who, What, How, When, and Where," relate to one another. Each question corresponds to one section of this survey, indicated by the numbered tag, near each question title. Each section lists its major subsections discussed in the survey.

Figure. 2 (Hohman et al., 2018, p.2)

[4] [Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software]

1.3 Why the communication designer needed?

The brain finds it more comfortable to process information that comes in the form of images rather than words or numbers. The right hemisphere recognises shapes and colours. The left side of the brain organises information analytically and sequentially, which is more active when people read text or look at spreadsheets. It takes much effort to look at a spreadsheet, but the information presented visually can be grasped in seconds. “Information visualisation is becoming more than a set of tools, technologies and techniques for large data sets. It is emerging as a medium in its own right, with a wide range of expressive potential.”(Manovich, L., 2011, p. 36) [5]

Communication designers are presenting the issues of the public, using visualisation as “things” to communicate to the participatory process. (Schoffelen et al., 2015) However, the visualisations are just components in the entire project; they need to be placed in the right positions and presented in a specific order. The “frame” of the transmission process is the narrative. Through the use of rhetorical narrative, designers gradually lead the audience to the professional context and make the information from visual works more precise and useful.

For instance, in the previous project "Deepfake Lab" [6], Deepfake technology has been demonstrated to ordinary users, using a series of annotated videos to introduce the visitor to the complexity of the algorithm and how its functioning leaves traces on the output. Towards the end of the site, the discussion will evolve along with increased quality for Deepfake video samples.

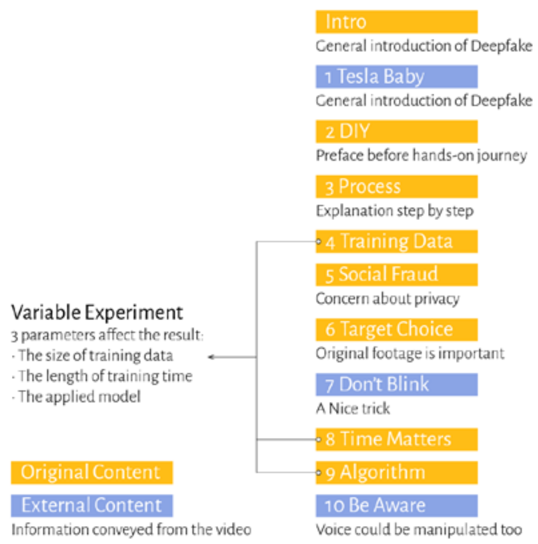


Figure. 3.1 Scheme of Deepfake Lab web experience

[5] Eric Rodenbeck [Stamen Design], in keynote lecture at Emerging Technology conference, March 2008

[6] Deepfake Lab

As shown in Figure 3.1 & 3.2, the first two videos *(Intro & 1 Tesla Baby)* draw viewers in by showing Deepfake videos with extremely low recognition difficulty, giving them a general introduction to this technology. The third video *(2 DIY)* is also easily identified as manipulated, a crude homemade Deepfake video that is a transition between the introduction and hands-on Journey.

The fourth video *(3 Process)* is like a catalogue. By showing the footage at different stages, the entire production of Deepfake video is divided into six parts. The fifth video *(4 Training Data)* shows an experiment of the size of the training data: How will different amounts of training images affect the final effect when other conditions are the same?

After revealing the importance of the training data size, we grabbed all the images of a colleague from her Facebook account. We provided the algorithm with these pictures as training data. So in the sixth video *(5 Social Fraud)*, we got Palame with 12 teeth all the time.

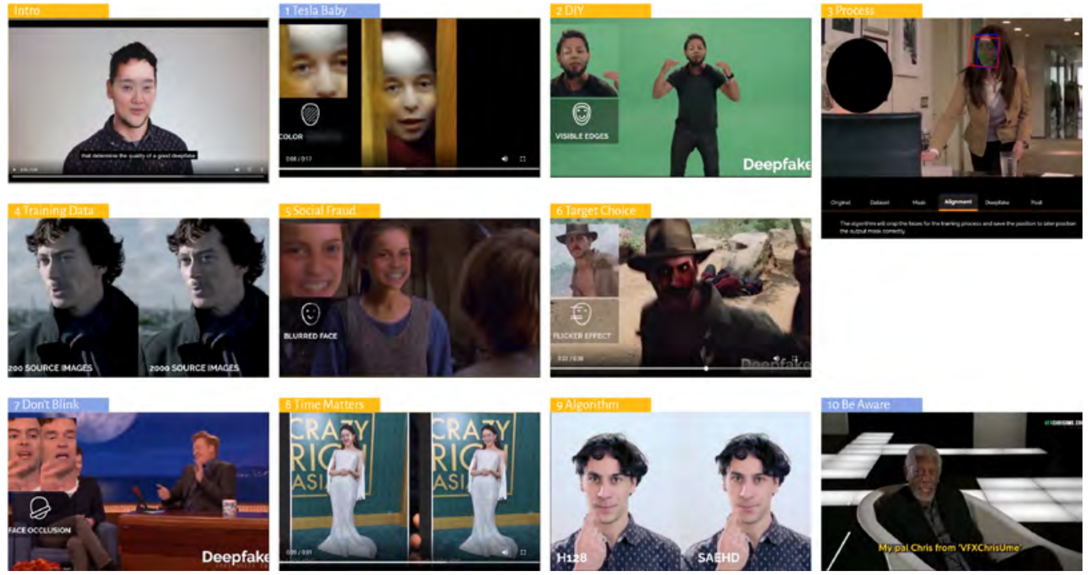


Figure. 3.2 Scheme of Deepfake Lab web experience

The next two videos *(6 Target Choice & 7 Don't Blink)* disclose that the choice of original footage is also essential. Subsequent experiments on the two control variables allowed the audience to understand how different training durations and different algorithm models affect the outcome. *(8 Time Matters & 9 Algorithm)*

The closing video *(10 Be Aware)* is well-crafted, alerting the audience to the maturity of this technology by showing a famous Hollywood actor saying strange things in his own voice.

1.4 The research path

As shown in Figure. 4, this research mainly consists of three parts:

1. Carrying out the preliminary investigation;
2. Learning about Computer Science expertise;
3. Applying interdisciplinary experiences to the design project

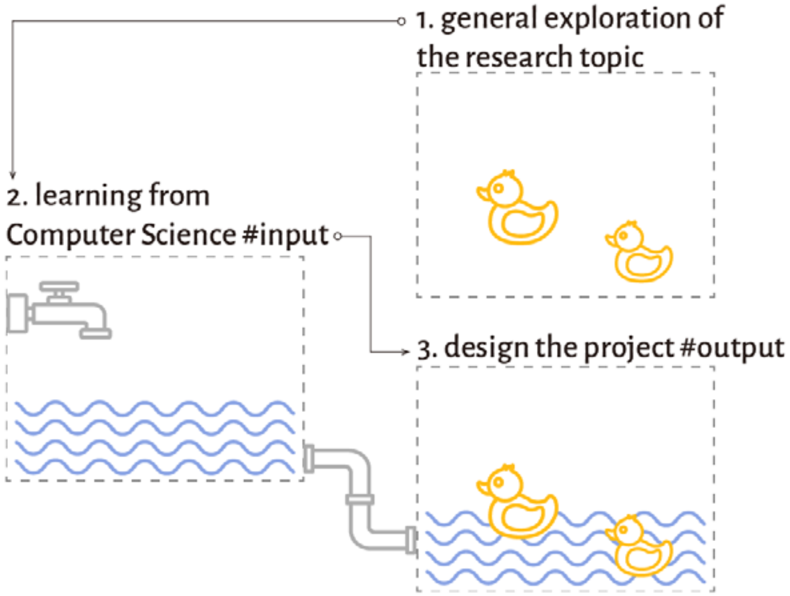


Figure. 4 Overview of the research path

The ducks in the diagrams symbolise the initial research and the theoretical experience that has been more in favour of design methodology since the master study career.

The water storage through the tap is an essential step of inputting information in the research, and the ripples represent a necessary element: expertise from the computer science field. In the end, the input of scientific knowledge and design experience has been combined to output the design work.

1. Carrying out the preliminary investigation
The original goal was to explain Machine Learning techniques to the public. After an experimental investigation of Machine Learning systems through popular scien media [7] and the exploratory literal review [8], the research topics and question framed, namely *how to explain ConvNet to the public as a communication designer*.

2. Learning about Computer Science expertise
Now it is a very novel stage. Being an amateur with no background in computer science, ConvNet has been learnt from scratch. In general, the main channels for understanding the algorithm are: published papers, online videos and articles (from Popular Science News as well as professional forums) and practically run the algorithm. Once learned some notions about convNet, visualisations to summarise, some of the key findings have been used.

[7] Machine Learning for everyone/ Deep Learning glossary/ Distill/ Reddit: machine learning
[8] Overview of Explainable Deep Learning list

3. Applying interdisciplinary experiences to the design project

At this stage, while refining the Deep Learning knowledge, the great challenge to balance the avoidance of overly technical narration and the clear conveyance of the expression, regarding the lay user, has been faced. Following the dimensionless “translation”, the author combines expertise from both communication design and computer science to execute the project.

LEARN CONVNET

AS AN AMATEUR

2.0 Learning path introduction

Since this is interdisciplinary research, the starting point is to comprehend the research object ConvNet through Computer Science.

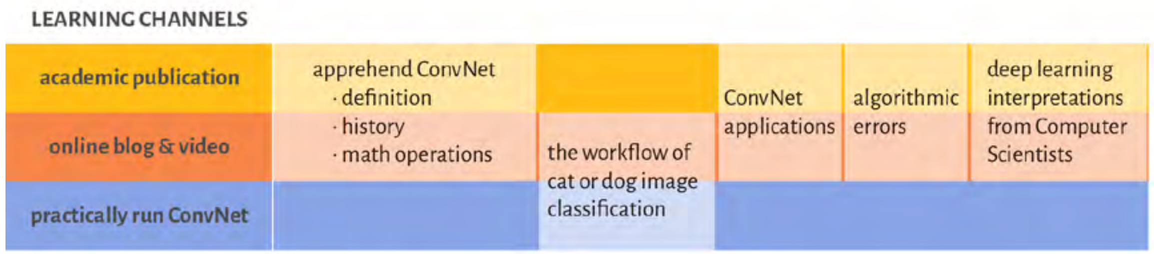


Figure. 5 Overview of the study path

As shown in Figure. 5, the three channels for acquiring Computer Science knowledge are not parallel; they often intertwine. For instance, it is common to find links to related research in some scientific and technological reports so that we can get the corresponding paper; the author also queries some topics from publications in search engines to understand how the society reacts. From figure 6.1-6.4, there are four kinds of research protocols the author used while learning ConvNet regarding different topics from professional discipline.

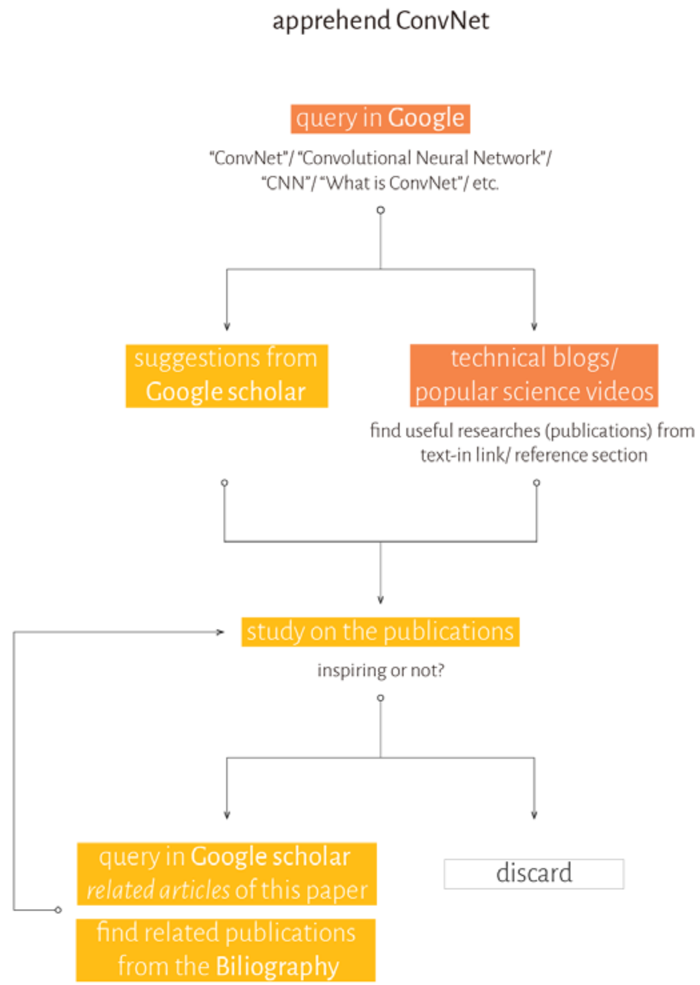


Figure. 6.1 Research protocol(1/4)

For instance, at the very beginning of the study, the author obtained a preliminary understanding of ConvNet through Google search, as shown in Figure 6.1. However, the highly cited academic papers are still the primary reference basis.

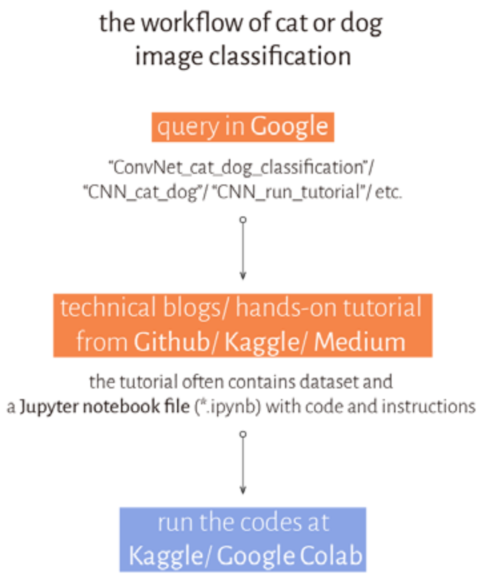


Figure. 6.1 Research protocol(2/4)

The practice is the best teacher. After learning the basic concepts, the algorithm has been run by referring to some simple tutorials on the web (recognising cat and dog images). Usually, we can find a wealth of learning resources from websites such as Github, Kaggle and Medium, which include training images and code with detailed instructions. Through this process, we can simply understand the basic workflow of classifying pictures by ConvNet.

After grasping the core value of the theoretical and practical aspects of ConvNet, it is time to expand to the edge, investigating the application of ConvNet and the error reporting. This step also used published academic

papers as the principal reference. However, given that the audience of this project is the general public, it is necessary to search through Google to verify whether the concerns about the problematic status quo from the academic world or the new applications can resonate with the public.

Figure. 6.1 Research protocol(3/4)

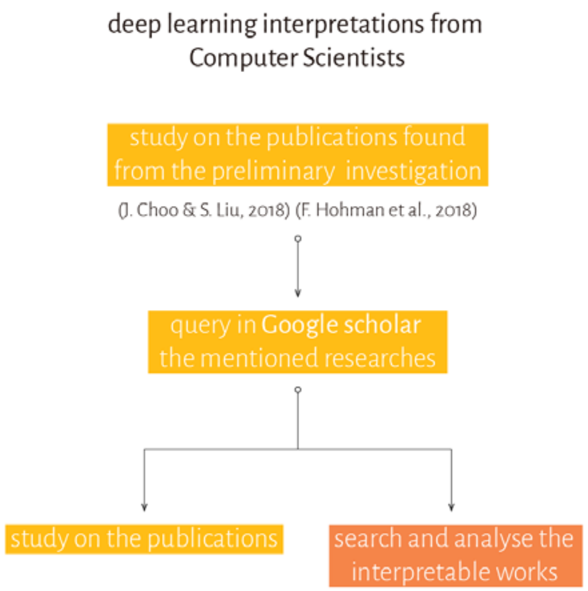


Figure. 6.1 Research protocol(4/4)

The tone of the search conducted in this step is to provide a literature review of the existing interpretable Machine Learning project. Following the two summary papers cited in the previous Chapter, (1.2 Summarise the visual interpretations of Deep Learning Techniques) by analysing the interpretation works mentioned in them and using the “related papers” function of Google Scholar, a thorough investigation has been made.

2.1 Theoretical study: Learn from Computer Scientists and Mathematician

2.1.1 Overview of Convolutional Neural Network history

Before 1970

The full name of ConvNet is Convolutional Neural Network, sounds like something taken from our brain. ConvNet is a bionic technique. Suppose we trace the origin of this fancy. In that case, it is the experiment done by biologists David H. Hubel and Torsten N. Wiesel in 1959, which greatly expanded the scientific community's knowledge of sensory signal processing.

The biologists inserted glass-coated tungsten wire micro-electrodes into the primary visual cortex of the anaesthetised cat. They then projected a light strip onto a screen placed in front of the cat, changing the spatial orientation of the Light Strip and recording the firing of neurons with micro-electrodes. They found that the emission was strongest when the band was at a particular spatial angle. Moreover, different neurons have different preferences for different spatial orientations. Some neurons have different response patterns to the bright and dark light zones.

These studies have shown people how the visual system presents simple visual features in the visual cortex and inspired computer scientists.

1970-1990

Japanese Computer Scientist Kunihiko Fukushima proposed “Neocognitron” in 1975, which is a hierarchical, multilayered artificial neural network. Neocognitron has played a role in connecting the past and the future. It was inspired by the model proposed by Hubel & Wiesel in 1959 and also served as an inspiration of Convolutional Neural Networks.

Inside the multilayered network, the input signal is transmitted forward until the output error occurs, and back-propagate the error information to update the weight matrix. Backpropagation is a primary algorithm for performing gradient descent on neural networks. First, the output values of each node are calculated (and cached) in a forward pass. Then, the partial derivative of the error concerning each parameter is calculated in a backward pass through the graph[9]. Another essential paper from D. E. Rumelhart et al. in 1985 showed through experiments that such networks could learn useful internal representations of data.

[9] Google Machine Learning Glossary

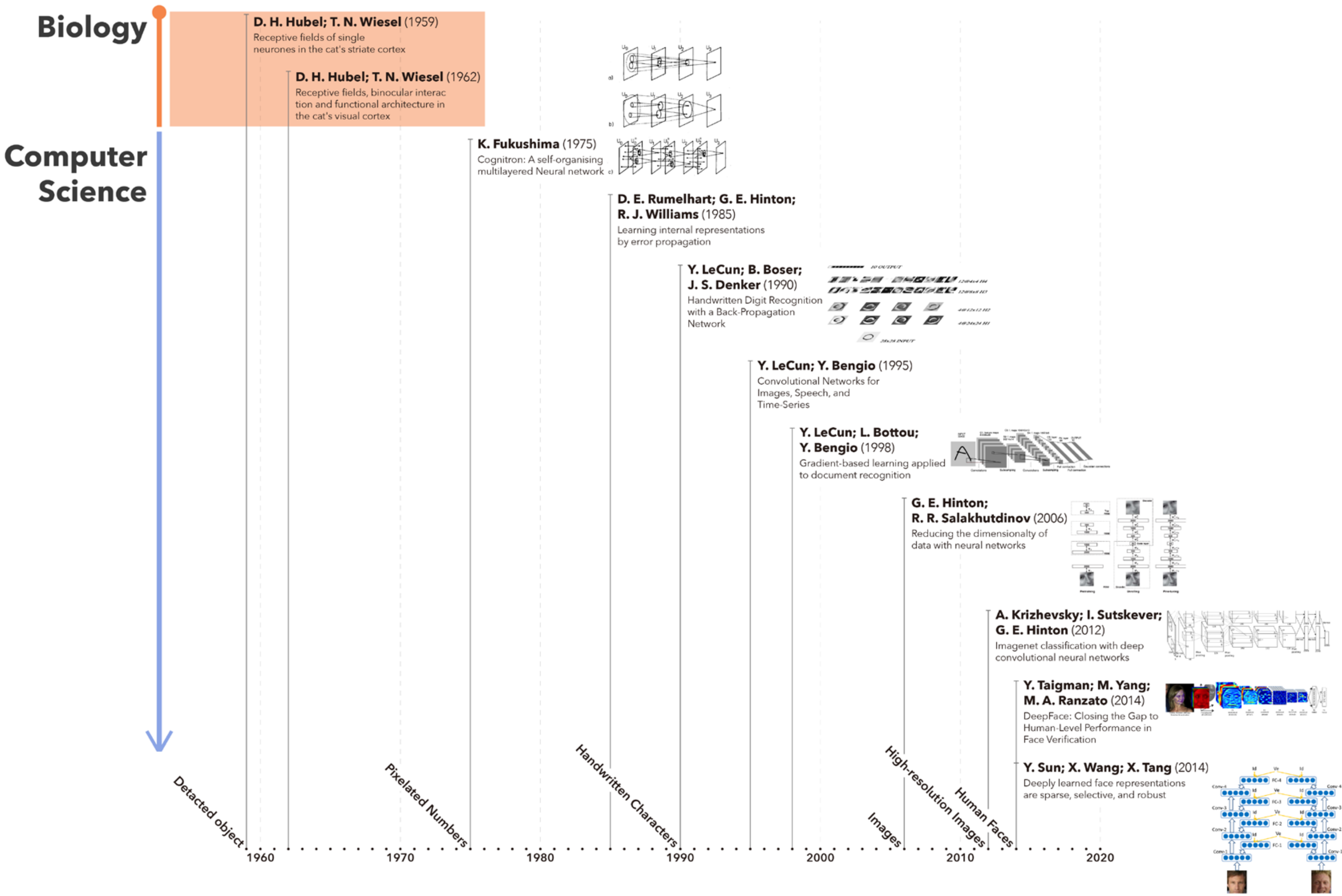


Figure. 7 The history of ConvNet

1990-2000

In these ten years, LeCun got the baton from Rumelhart and applied Back-Propagation algorithm while designing the CNN model for recognising handwritten digits, documents, Speech, and Time-Series. In 1998 the proposal of the LeNet-5 model marked the official formation of ConvNet.

2000-2010

However, over the next right years, there was no breakthrough with ConvNet technique because the researchers realised that the amount of computation required for Back-Propagation training in multilayer neural networks was so large that the hardware computing power at that time was utterly impossible to achieve.

In 2006, G. H. Hinton pointed out that multi-hidden-layer neural networks have better feature learning ability and their training complexity can be effectively alleviated through layer-by-layer initialisation. In the context of GPU-accelerated hardware and the widespread application of Big Data recognition, ConvNet takes off again.

2010-2020

During this period, the research and development of ConvNet are in full swing, and academic literature is booming. The excellent performance of AlexNet in the ImageNet competition in 2012 helped establish the critical position of ConvNet.

After ConvNet became an instant hit in 2012, its application field is no longer limited to handwritten digit recognition and voice recognition. Face recognition has become one of its important application areas. During this time, DeepFace and DeepID, as two relatively successful high-performance face recognition and authentication models, became ConvNet's iconic research results in the field of face recognition. In an attempt to theoretically explain ConvNet's powerful feature extraction capabilities, DeepID, whose team is led by Professor Tang, has analysed the internal structure of the network.

In 2016, ConvNet once again gave people a surprise: the intelligent robot AlphaGo developed by Google based on deep neural networks and search trees defeated humans in Go.

2.1.2 Large Scale Visual Recognition Challenge and classic models ILSVRC

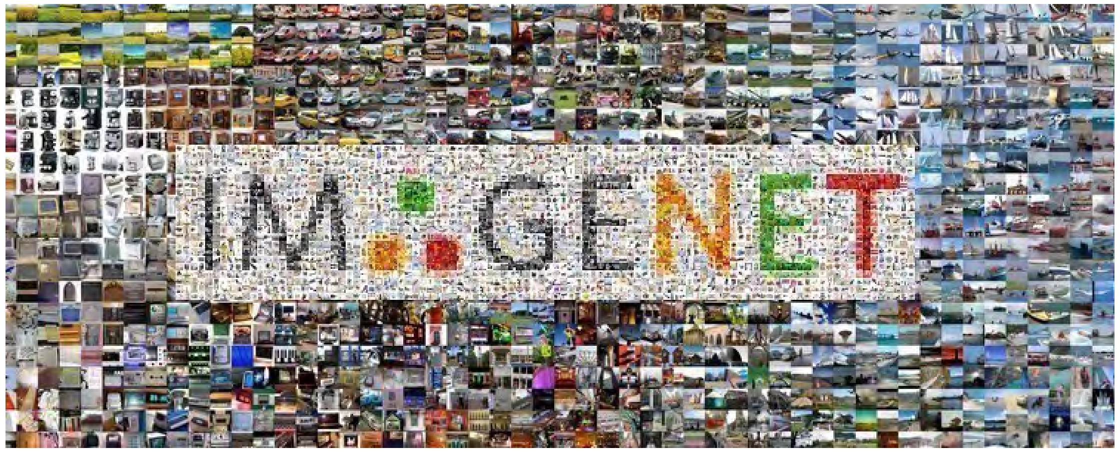


Figure. 8 ImageNet *(source)*

ILSVRC

Needless to say that ConvNet is useful. But how do we evaluate a ConvNet model? Computer scientists often use the Error Rate of ConvNet to classify objects to assess the performance. The lower Error Rate, the better the ConvNet model is.

ImageNet is an extensive visual database designed for use in visual object recognition software research, and ILSVRC (ImageNet Large Scale Visual Recognition Challenge) evaluates algorithms for object detection and image classification at a massive scale. The challenge allowed researchers to compare progress in detection across a wider variety of objects and to measure the progress of computer vision for large scale image indexing for retrieval and annotation.

From the figure. 9, we can draw the following insights:

1. By comparing the results around 2012, the Error Rates of classifying objects using other algorithms in the early stage was significantly higher than that using ConvNet to execute classification tasks.

2. Since 2013, ConvNet has grown exponentially, and the Error Rate of the Champion model has dropped dramatically. Even since 2015, the performance of ConvNet object classification has outperformed humans.

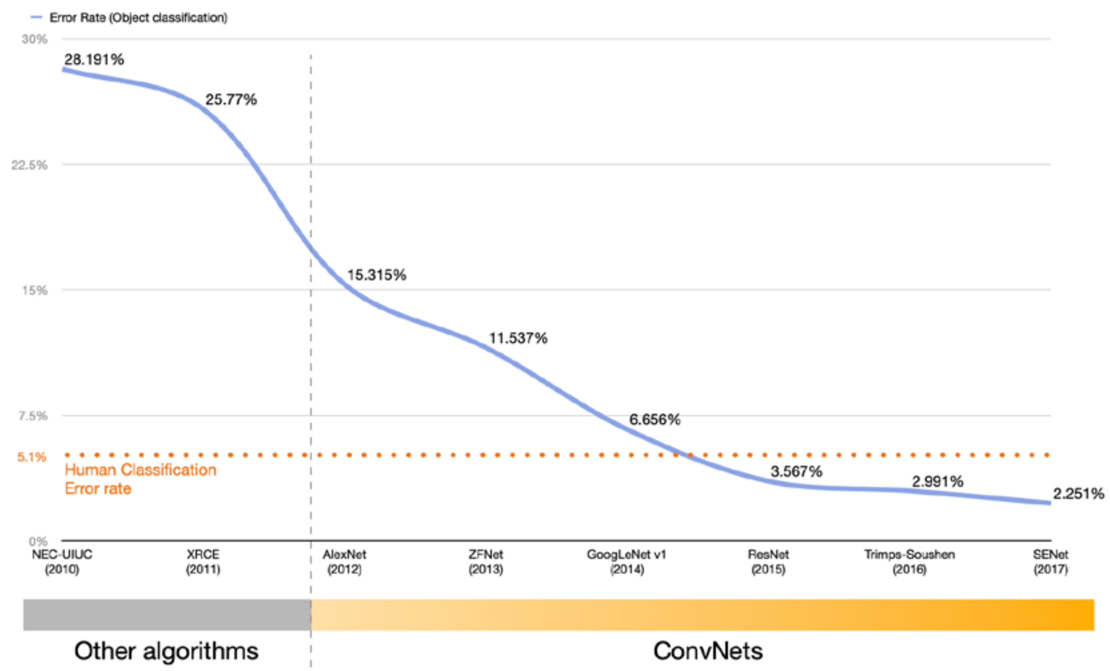


Figure. 9 LSVRC best object classification results from 2010-2017

(Data from ImageNet)

Classic Model

Deep learning researchers have always come up with new ideas and goals to design different ConvNet models to meet the needs of ever-changing applications. Through the parallel comparison of the following four classic models, we can more or less get a glimpse of the development trend of the ConvNet model structure.

It is clear from the table that the number of layers of the model has increased over the last 20 years, indicating that the algorithm has become more complex while relying on higher computational power to process more complex recognition objects.

Combining with the architecture images, we can find that the development of ConvNet is not a linear superimposing the layers, pursuing depth, but simultaneously expanding vertically.

	LeNet	AlexNet	GoogLeNet	ResNet
release year	1998	2012	2014	2015
layers	7	8	22	34
recognition object	handwritten digit	Image	Image	Image
architecture	Figure 10.1	Figure 10.2	Figure 10.3	Figure 10.4

Table. 1 Classic ConvNet models

For instance, GoogLeNet made efficient use of the internal computing of the network, increasing the depth and width of the network while ensuring the stability of computing consumption.

By observing the visual representation of LeNet-5, the earliest ConvNet model, we could find that it consists of convolution, subsampling(also pooling refer to other figures) and fully-connection operations. These operations are visually presented as hierarchical layers in the model.

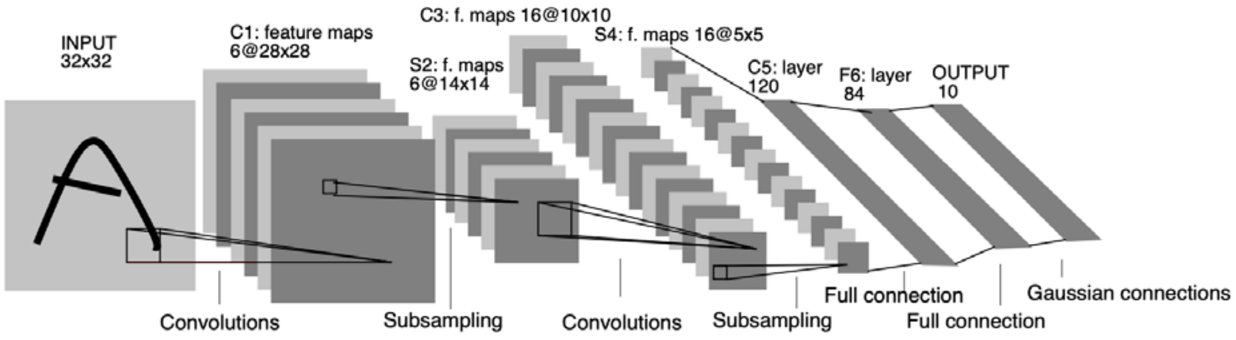


Figure. 10.1 LeNet (1998)

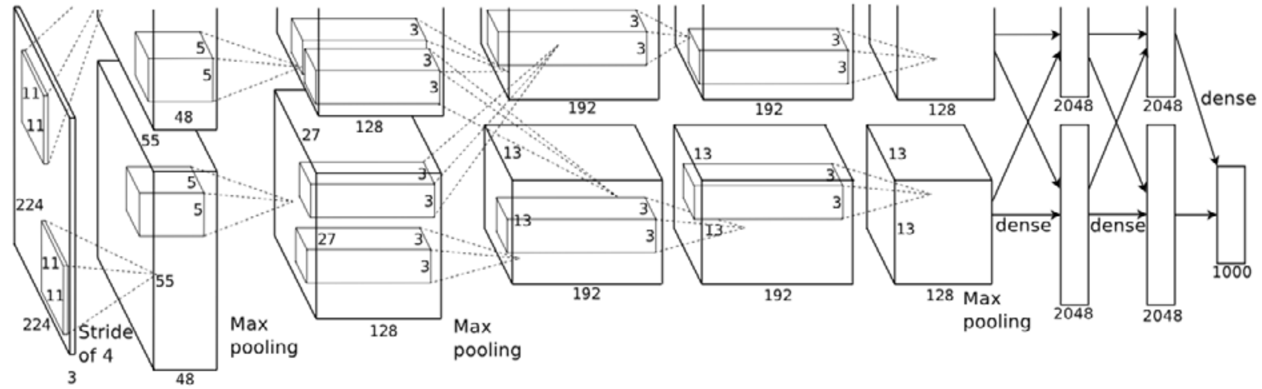


Figure. 10.2 AlexNet (2012)

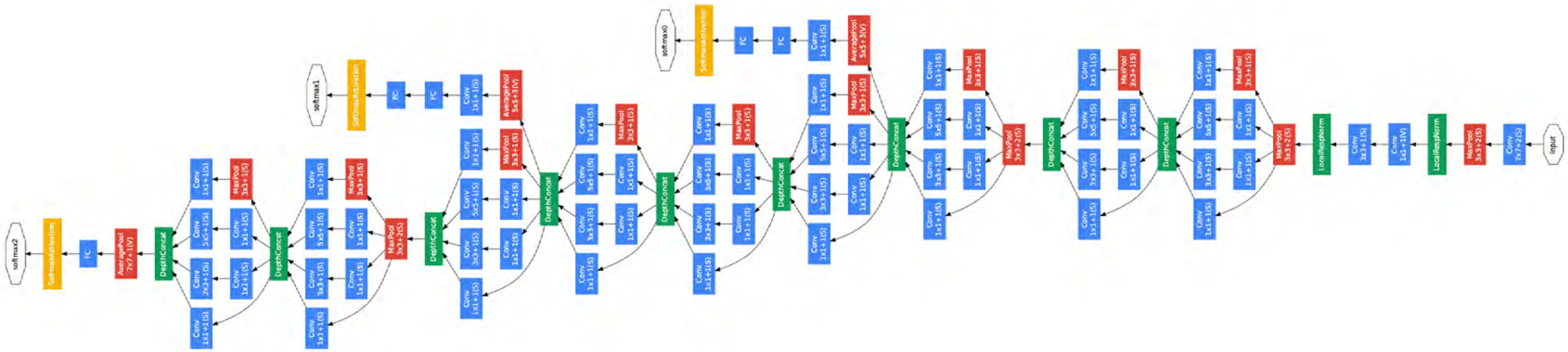


Figure. 10.3 GoogLeNet (2014)

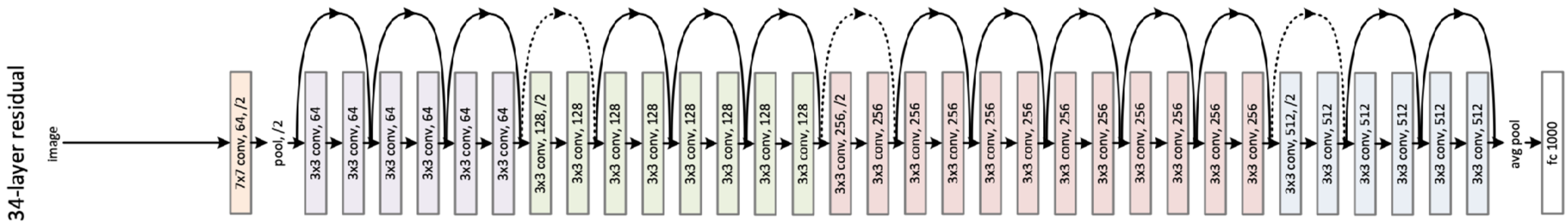


Figure. 10.4 ResNet (2015)

2.1.3 The applications relate to Convolutional Neural Network

Understanding its application field is an essential means to comprehend a new technology from the side: Get rid of the complex and trivial model parameters, start from the vivid use scenarios.

As shown in Figure 11, ConvNets are mainly used in Computer Vision and Natural Language Processing. Beyond driverless technology for the near future, ConvNet has come to us, hidden in mobile phones, hidden in intelligent virtual assistants.

In daily life, ConvNet silently plays an important role: quickly recognise faces when unlocking the phone, converts users' voice commands from sound waves to text, and intelligently classify photo albums for easy retrieval. In subdivided professional fields, ConvNet also saves a lot of time and labour costs with outstanding performance: security check, radiology, botany and zoology. Singapore National Parks use ConvNet to help collect bird population because not everyone can correctly identify birds.

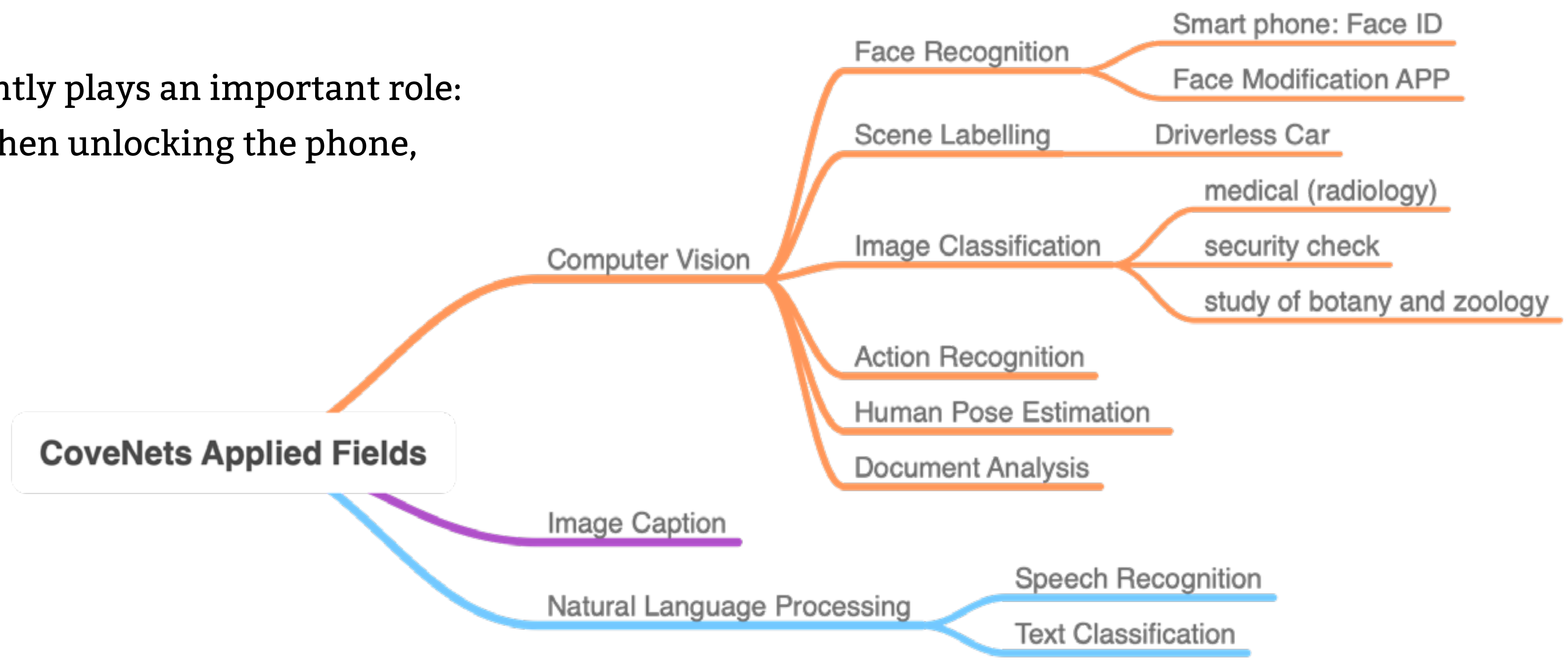


Figure. 11 ConvNets applied fields

information from (Bhandare et al., 2016)

AlphaGo

It has long been thought that Go is the most challenging classic game for artificial intelligence because of its vast search space, the difficulty of evaluating a game and evaluating the location of the game. Researchers from Google DeepMind have introduced a new method to the computer Go program, combines Monte Carlo Tree Search with the deep Convolutional Neural Network, which achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by five games to zero, a feat previously thought to be at least a decade away.

Monte Carlo Tree Search uses the Monte Carlo rollouts to estimate the value of each state in the search tree. Researchers used images from 19 × 19 to represent the positions of the Go board and used convolution layers to construct a neural network representation of the Go positions.

They used ConvNet to effectively reduce the depth and width of the search tree, by evaluating the position of a chess player through a valuation network, and by sampling strategy networks to select the move

2.1.4 Algorithmic errors

Although ConvNet outperforms humans in some respects, the technique still has limitations today. The most well-known case was in 2015, Google made the mistake of labelling a black girl as a gorilla when it applied Machine Learning technique to classify images, which caused widespread concern in society. However, two years later, Google still failed to tackle the issue properly. They simply removed the terms "gorilla" and "chimpanzee" to ensure that no similar errors would appear, but correspondingly, real chimpanzees unable to be marked and retrieved[10].

Google's cautiousness towards gorilla images illustrates the shortcomings of existing Machine Learning technology. With enough data and computing power, we can train the software to categorise images or transcribe speech to high accuracy. However, this cannot easily exceed the training process. Even the best algorithm cannot use common sense or abstract concepts to refine the interpretation of the world like humans.

[10] When It Comes to Gorillas, Google Photos Remains Blind

How do these algorithmic errors come about? Let us go through the whole workflow of ConvNet first.

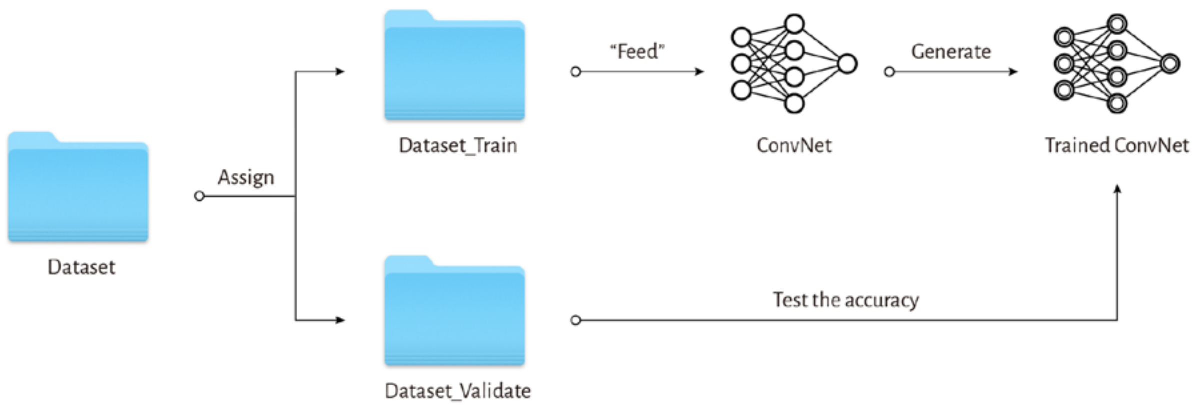


Figure. 12.1 ConvNets workflow

· At the beginning, we have a folder which contains many data; in this case, the data are pictures. Typically, researchers divide these data into training and validation, where the training data is annotated, and the validating data is not.

· At this point, the ConvNet is just a frame, consisting of code, which cannot yet recognise images. Then we use the data from "Dataset_Train" to feed the algorithm(ConvNet in this case). Like educating students at school, the training data as the textbook.

· Afterwards, ConvNet begins to analyse the annotated pictures layer by layer, according to

the architecture designed by the researchers, find and integrate the detailed features in a digital matrix. Then ConvNet associates the annotation information of the picture with the matrix and saves it in the model[11].

· Therefore, we now have a trained ConvNet, that is, it can classify objects(the range depends on what we feed during the training process). Next, the model needs to be validated. It analyses pictures again, but these pictures come from the validate folder and have not been labelled. The analysis process is similar to the training process in the previous step; the algorithm generates a string of code for each new picture. The only difference is the last step: these newly generated codes are compared repeatedly with the codes stored in the training phase, and every result is a numerical representation of the probability that the prediction has been accurate, the closer the number gets to one, the better the classification.

· In the end, the researchers analyse the results of the validation tests to further debug the model structure, then repeat the same routine.

[11] For example, if the goal of this task is to make the machine classify five kinds of fruits, apples, bananas, oranges, avocados and grapes. So in this step, we need to let the algorithm analyse many annotated pictures of these five fruits. After a series of calculations, the algorithm generates five strings of different codes to represent the five fruits, just like different products in the supermarket are labeled with different bar codes.

The whole process seems rational and well-tested, so where is the problem. As figure 11.2 shown, many of the steps of human intervention. Such as the collection and annotation of the dataset and the design of the recognition model.

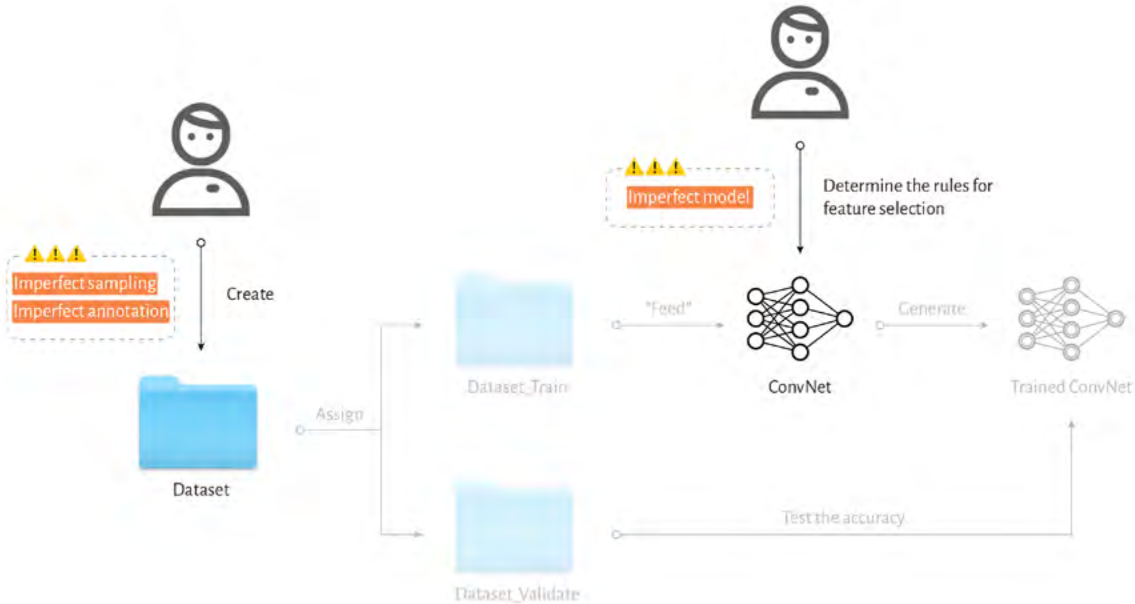


Figure. 12.2 ConvNets workflow

Imperfect sampling _ the causes of error (1/3)

Datasets are the basis of Machine Learning. If the dataset itself lacks representativeness, it cannot objectively reflect the reality, and algorithmic decision-making will inevitably be unfair. The typical manifestation of this problem is a mismatch. Due to the convenience of data collection, data sets tend to be more "mainstream" and accessible groups, and thus uneven distribution at the level of race and gender.

Shankar et al. suggested that examining the geo-diversity of open datasets is critical before adopting a data set for use cases in the developing world. In the course of their research, they analysed two popular public datasets: ImageNet and Open Images, which, while not covering all possible use cases, are commonly used standard datasets for most people.

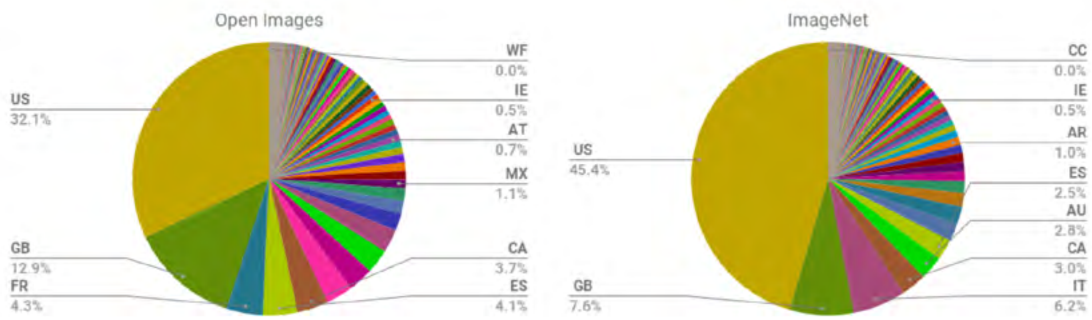


Figure. 13 Distribution of two popular public datasets
(see Figure 1 of Shankar, S et al., 2017, No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World)

In both datasets, top represented locations include the US and Great Britain. Most of the sample data were from the US and countries across North America and Europe. Meanwhile, China and India – the two most populous countries in the world – were represented with only 1% and 2% of the images, respectively. The finding explains why the traditional wedding dress Hyderabad raters are likely to be misclassified as chain mail, a kind of armour, by a classifier trained on the ImageNet data.

The standard open-source datasets do not have sufficient geographic diversity to represent most developing countries broadly. However, this research inspired us to evaluate the appropriateness of a given dataset before using and the importance of innovative datasets.

Improperly sampled datasets exist not only geographically, but also in darker-skinned and lighter-skinned. Research has found that the high error rate of AI recognition for dark-skinned females may explain some of the error[12]. The paper, ”Gender Shades: Intersectional accuracy disparities in commercial gender classification”, evaluated three commercial gender classification systems from different countries in parallel and found that the highest error rate in identifying lighter-skinned males was only 0.8 per cent, while the recognition error rate of darker-skinned females was as high as 34.7 per cent. The researchers then analysed the distribution of face datasets used in the test, as shown in the Figure. 14.

Among these papers, Adience and IJB-A are well-known datasets for facial recognition, and PPB created by the original author. By observing the first two datasets,

[12] Facial recognition software easily IDs white men, but error rates soar for black women

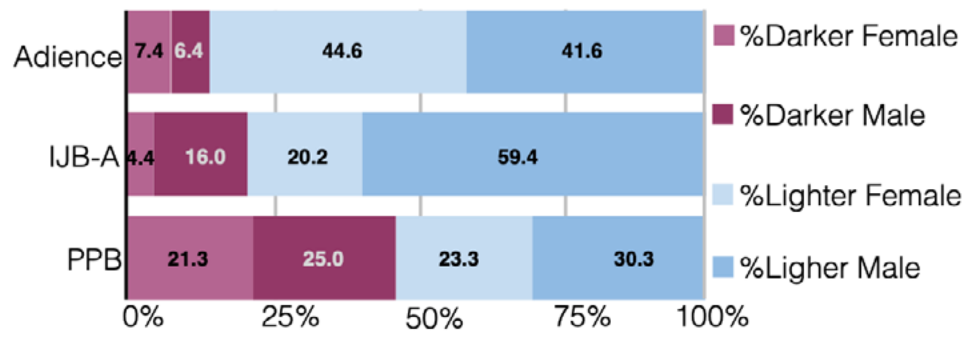


Figure. 14 Distribution of three datasets for face recognition
(see Figure 3 of Buolamwini, J., & Gebru, T., 2018, Gender shades: Intersectional accuracy disparities in commercial gender classification)

we find that lighter-skinned males have a much higher proportion of data than the others, whereas the data about darker-skinned females have the lowest percentage. Although the companies behind the software in the test did not disclose the dataset they used to train the algorithm, the significant difference in error rates between lighter-skinned males and darker-skinned females is always worth thinking.

Indeed, from the perspective of human genes and face recognition technology, the machine needs to preprocess and extract features of the face image during the process of face detection, analysis and recognition. Therefore, the darker the skin tone, the more difficult it is to extract facial features, especially in some bad lighting conditions, it is much more difficult to detect and distinguish.

On the other hand, face recognition relies on data training to a great extent, and darker-skinned subjects have less training data in the whole industry, so the face recognition results are unsatisfactory[13]. Amazon found that the reason for the deviation of its recruitment system is that the raw dataset used by the algorithm is the company's past employee data-in the past, Amazon hired more men in the past years. The algorithm has learned this feature of the dataset, so it is easier to ignore female job applicants in decision-making.

In this situation, the existing social biases are brought into the dataset. When the raw data is the result of social discrimination, the algorithm will also learn and amplify the bias.

[13] 人脸识别有歧视？ 识别准确率黑人比白人差很多，厂商们这样回应 (Facial recognition is discriminatory? Black people are much less accurate than white people, the manufacturers responded)

Imperfect annotation _ the causes of error (2/3)

It is no doubt that the objectivity of the dataset is crucial. For some unstructured data sets (such as a large number of descriptive texts, pictures and videos), the algorithm can not directly analyse them. At this point, it is necessary to manually annotate the data to extract structured dimensions for training algorithms. To give a straightforward example, sometimes Google Photos will ask you to help determine whether a picture is a cat, and then you will participate in the marking of that picture.

When an annotator is asked a “cat or dog” question, the worst result is just a wrong answer; but when asked a “beauty or ugliness” question, a bias is created. As a data processor, the annotator is often asked to make some subjective value judgments, which in turn becomes a significant source of bias.

ImageNet is a case in point: as the world's largest image recognition dataset, many of the pictures on the website are manually annotated and labelled with various subdivisions. "Although we cannot know if the annotators have such prejudices. But they define what "losers", "sluts" and "criminals" should look like. The same problem may also occur in seemingly On the label of 'harmless'. After all, even the definition of 'man' and 'woman' is open to question."

Trevor Paglen is one of the founders of the ImageNet Roulette[14] project, which aims to show how opinions, prejudices and even offensive attitudes affect artificial intelligence. “The way we label images is a product of our worldview, ” he said. “Any classification system will reflect the values of the classifiers. ” Under different cultural backgrounds, people have prejudices against other cultures and races.

The annotating process precisely transfers personal bias to the data, which is absorbed by the algorithm, thus generating a biased model. Nowadays, manual annotating service has become a typical business model, and many technology companies outsource their massive data for marking. It means that algorithmic bias is spreading and amplifying through the process of invisibility and legitimation.

[14] Excavating AI- The Politics of Images in Machine Learning Training Sets

Imperfect ConvNet model (feature selection) _ the causes of error (3/3)

Another bias comes directly from humans. The algorithmic engineers, who are predominantly male, come from a specific ethnic demographic, have grown up in areas of high socioeconomic status, and are mostly people without disabilities. They are a relatively homogeneous population, so trying to think broadly is a challenge.

The engineers are involved in the entire system from beginning to end, including goal setting for Machine Learning, which model to use, which features to select (data labels), data preprocessing, and so on. Inappropriate goal setting may introduce biases from the beginning, such as the intent to identify criminals by their faces; however, more specific personal preferences are presented in the selection of data features.

I do not think there is genuinely an impartial person, so whether we are building an AI system is fair or not is open to question. But of course, we can do much better than what we do.

2.1.5 Visual interpretations made by Computer Science experts

As we discussed before, the State-of-the-art ConvNet technique has improved significantly in recent years, reaching levels of superiority over humans, especially in object recognition, (Russakovsky et al., 2015), and humans rely on artificial intelligence techniques more than ever. Even with unprecedented advances, the lack of explanation of Deep Learning model decisions and control over their internal processes remain major flaws in crucial decision-making processes, such as precision medicine and law enforcement. In response, efforts are underway to enable humans to interpret and control Deep Learning. (Choo, J., & Liu, S., 2018)

As a non-professions, the academic review is needed for understanding how the Machine Learning decision-making process is visualised scientifically, as well as to be inspired for a general but accurate method to explain ConvNet visually. As Table. 2 is shown, there are 18 visual analytic works collected, among them, more than ten interpretations are presented in the form of a webpage, and 4 cases are developed as toolkits that require local installation. Due to the lack of programming foundation, the author will focus on the interactive interpretation websites in this research.

(CNN/ DL/ ML) Interpreting Visualisation by Computer Science

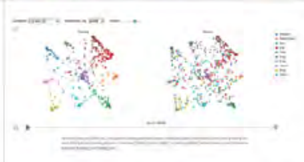
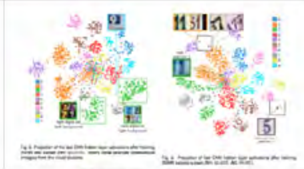
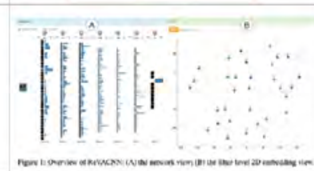
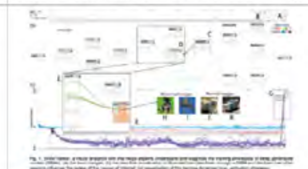
Year	Title of paper	Arthor	Method	interactive details	Goal	Topic	Snapshot	Link	Comments
2015	An Interactive Node-Link Visualization of Convolutional Neural Networks	A. W. Harley	interactive website	User provide input data	understanding	CNN		https://www.cs.ryerson.ca/~aharley/vis/	
2020	CNN EXPLAINER: Learning Convolutional Neural Networks with Interactive Visualization	Z. J. Wang, R. Turko, O. Shaikh, H. Park, N. Das, F. Hohman, M. Kahng, and D. H. Chau	interactive website	User could switch between overview and restricted details	understanding	CNN		https://poloclub.github.io/cnn-explainer/	
2016	Towards Better Analysis of Deep Convolutional Neural Networks	M. Liu, J. Shi, Z. Li, C. Li, J. Zhu, S. Liu	interactive website	User could tune parameters	understanding	CNN		http://shixialiu.com/publications/cnnvis/demo/	
2020	Visualizing Neural Networks with the Grand Tour	M. Li, Z. Zhao, C. Scheidegger	interactive website	User could watch a live training process.		CNN		https://distill.pub/2020/grand-tour/	
2017	Visualizing the Hidden Activity of Artificial Neural Networks	P. E. Rauber, S. G. Fadel, A. X. Falca˜o, and A. C. Telea	local toolkit (I am not able to install it, but I put a figure inside the paper)			CNN		http://www.cs.rug.nl/svcg/People/PauloEduardoRaube	
2018	The Building Blocks of Interpretability	C. Olah, A. Satyanarayan, I. Johnson, S. Carter, L. Schubert, K. Ye and A. Mordvintsev	interactive website	User could tune parameters		CNN		https://distill.pub/2018/building-blocks/	
2014	(no publication) Computer Science PhD student, Stanford University	Andrej Karpathy	interactive website	User could watch a live training process.		CNN		https://cs.stanford.edu/people/karpathy/convnetjs/dem https://github.com/karpathy/convnetjs	
2016	ReVACNN: Steering Convolutional Neural Network via Real-Time Visual Analytics	S. Chung, C. Park1, S. Suh, K. Kang, J. Choo, B. C. Kwon	local toolkit	by inserting parameters the system will properly perform the training process		CNN			
2018	Analyzing the Training Processes of Deep Generative Models	M. Liu, J. Shi, K. Cao, J. Zhu, S. Liu	demo video			DGM		http://shixialiu.com/publications/dgmtracker/video.mp	

Table. 2a Academic reviews of visual interpretation from Computer Scientists

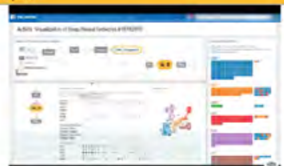
Year	Title of paper	Arthor	Method	interactive details	Goal	Topic	Snapshot	Link	Comments
2017	ACTIVIS: Visual Exploration of Industry-Scale Deep Neural Network Models	M. Kahng, P. Y. Andrews, A. Kalro, and D. H. Chau	demo video			DNN		https://minsuk.com/research/activis/	
2015	Understanding Neural Networks Through Deep Visualization	J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson	demo video			DNN		http://yosinski.com/deepvis#toolbox	
2017	DeepEyes: Progressive Visual Analytics for Designing Deep Neural Networks	N. Pezzotti, T. Holt, J. V. Gemert, B. PF. Lelieveldt, E. Eisemann	local toolkit			DNN			
2016	Embedding Projector: Interactive Visualization and Interpretation of Embeddings	D. Smilkov, N. Thorat, C. Nicholson, E. Reif, F. B. Viégas, M. Wattenberg	interactive website	User could tune parameters		Embedding		https://projector.tensorflow.org/	
2018	GAN Lab: Understanding Complex Deep Generative Models using Interactive Visual Experimentation	M. Kahng, N. Thorat, D. H. Chau, F. B. Viégas and M. Wattenberg	interactive website	User could watch a live training process.		GAN		https://poloclub.github.io/ganlab/	
2018	GANViz: A Visual Analytics Approach to Understand the Adversarial Game	J. Wang, L. Gou, H. Yang, H. Shen	demo video			GAN		notation video: https://www.youtube.com/watch?v=utz	
2017	Direct Manipulation Visualization of Deep Networks	D. Smilkov, S. Carter, D. Sculley, F. B. Viégas, M. Wattenberg	interactive website	User could tune parameters		neural networks		http://playground.tensorflow.org	
2017	Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow	K. Wongsuphasawat, D. Smilkov, J. Wexler, J. Wilson, D. Mane, D. Fritz, D. Krishnan, F. B. Vi ˙egas, and M. Wattenberg	embedded inside TensorFlow platform (the figure is from paper)	User could visualise their models		neural networks		https://github.com/tensorflow/tensorboard/blob/maste	
2016	LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks	H. Strobelt, S. Gehrmann, H. Pfister, A. M. Rush	interactive website	User could tune parameters		RNN		http://lstm.seas.harvard.edu/client/index.html	

Table. 2b Academic reviews of visual interpretation from Computer Scientists

Many interpreting visualisations use the website as the carrier because it allows the user to tune the parameters to show the dynamic changes of the visualisation works, thus achieving the role of interpretation. The ten websites mentioned in the table are extracted and analysed in depth from two aspects: layout and degree of interactivity, as shown in figure 15.

Website layout

Consider the content layout of the webpages, and the works can be divided into two categories: equipped sizable text instructions and those without plenty of text. As shown the cases on the left in Figure 15 (a, b, c, d, e, f), the text of these web pages is usually about the detailed analysis of the top visualisation, which aims to help audiences understand some basic terms and how the Machine Learning model works; the cases (g, h, i, j) on the right side are relatively simple, and the entire interface is only used to display the main visualisation.

Links

(a) Visualizing Neural Networks with the Grand Tour

(b) The building blocks of interpretability

(c) CNN EXPLAINER

(d) Gan Lab

(e) Direct-manipulation visualization of deep networks

(f) ConvNet JS CIFAR-10 demo

(g) LSTMVIS

(h) Embedding Projector

(i) An interactive node-link visualization of ConvNets

(j) CNN Vis

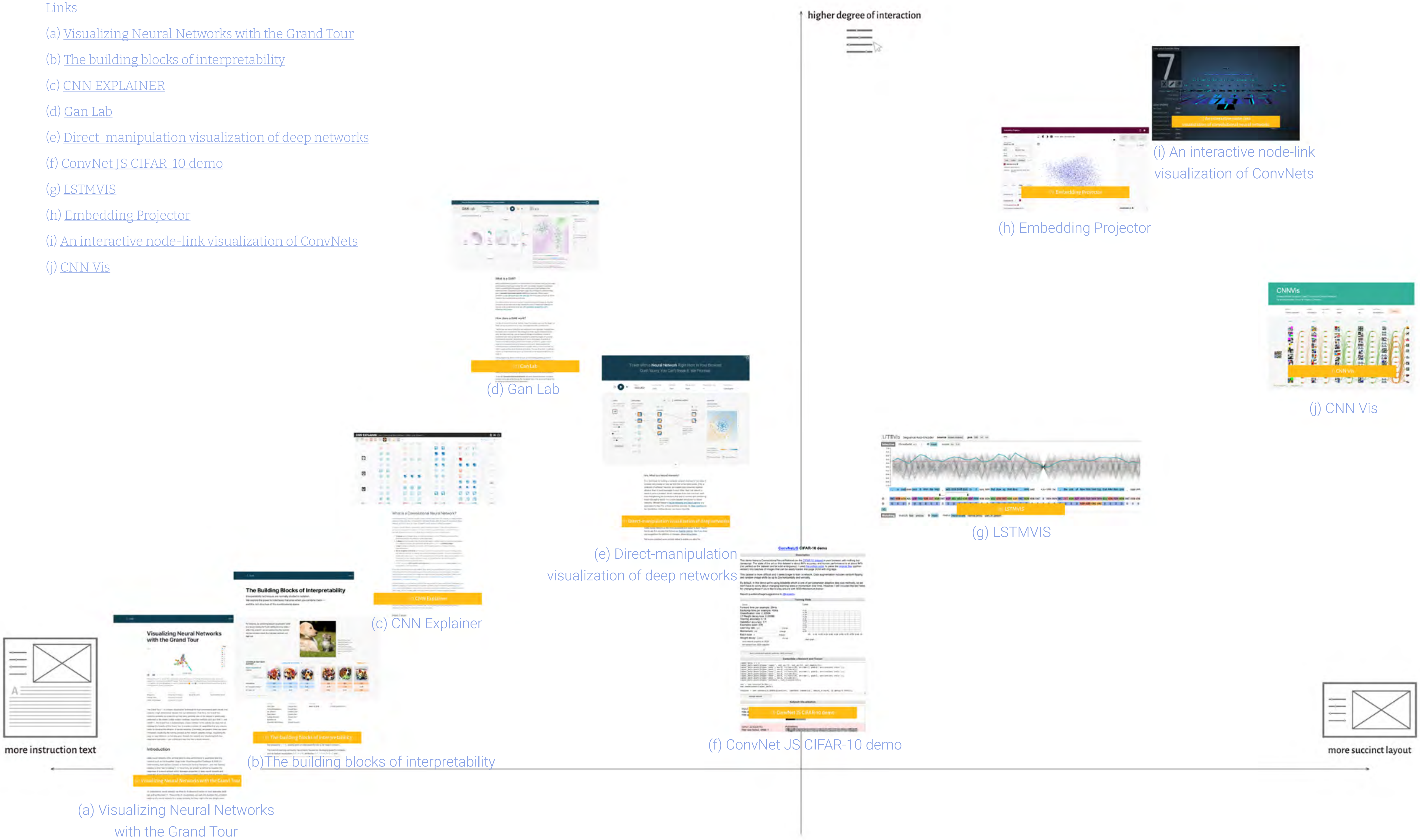


Figure. 15 Analysis of interactive Machine Learning interpreting website

Interpretation works like Visualizing Neural Networks with the Grand Tour(Li et al., 2020) and The Building Blocks of Interpretability (Olah et al., 2018)(See Figure. 15a,b) are published on Distill[15], which is a peer-reviewed scientific journal covering Machine Learning. It should be noted that this Journal website is following the "main visualisation + research dissertation" structure pattern. It can be seen that in both cases, the layout of the page is not a means used by the author to reinforce the interpretation. Similar examples are "Gan Lab" and "Direct-manipulation visualisation of deep networks"(Figure. 15d, e) whose content structures are similar because their principal researchers are from Google Brain.

The main visualisation of CNN EXPLAINER (Wang et al., 2020)(See Figure. 15c) reveals the working process of ConvNet by showing the images of each step, and the connections between each image disclose the complexity of the neural network. Indeed it is necessary to explain every column of images symbolises a layer inside the ConvNet model. The ConvNetJS Demo, developed by Andrej Karpathy, is unique in that the main

[15] Distill

visualisation does not show up at the top of the page, but instead requires users to continuously scroll the pages to see the long and narrow visualisation intact. Furthermore, based on the same object ConvNet, Harley got rid of the limited two-dimensional plane space and used the depth view to tile the image of each stage.

Degree of interactivity

Generally, these visualisations allow users to interact with the web page to a certain extent, including:

· dragging the slider to change the parameters (a, e)
· clicking on different images to switch input data (b, c, d, g)
· clicking the drop-down menu to select various parameters (e, h, j)
· entering a number to change the parameter (f),
· using the mouse to hover, click or drag the interface to highlight some local details (b, c, e, h, i, j)
· using the mouse to scribble on the blank canvas to create a real-time input image (i).

Visualisation types are chosen by CS experts

The selection of the visual model has a significant

relationship with the target object of interpretation. Among the ten interactive visualisations selected above, the objects interpreted are ConvNet(c, f, i, j), GAN(d), Neural Networks(a, e, h) and LSTM(g).
In the case of CNN, its distinctive feature is its strict hierarchical system, so the researchers used

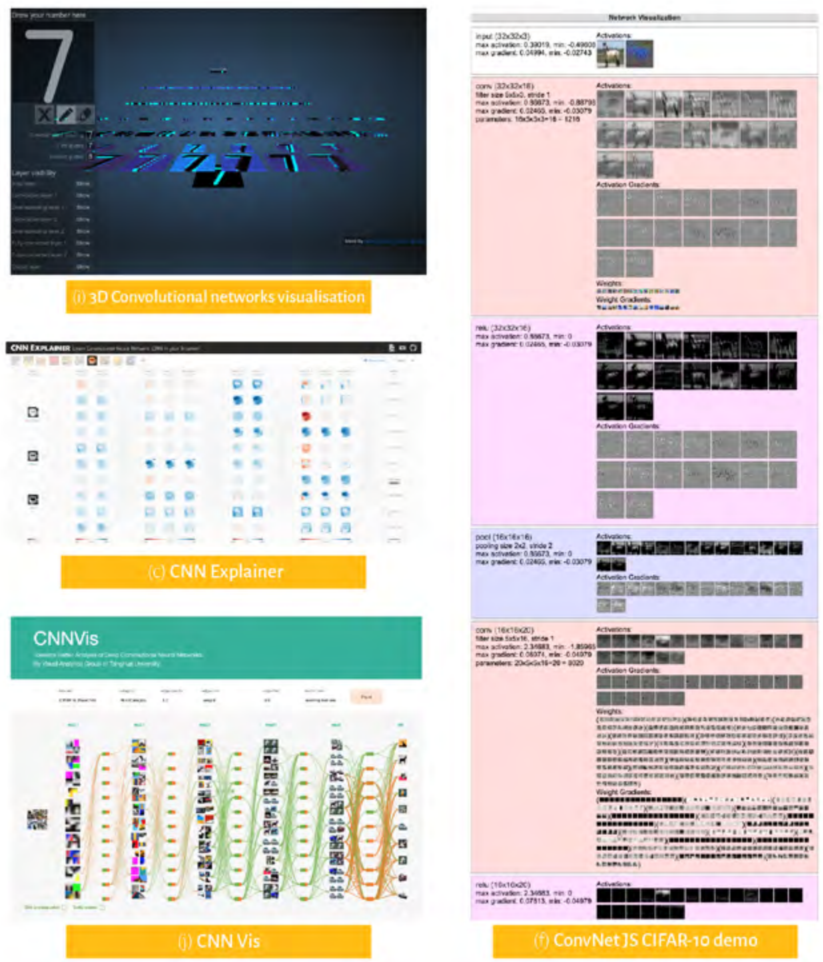


Figure. 16 Visualisations of tiling images

visualisation to interpret the decision-making process by tiling the states of the input data(image) after each operation. (see Figure 16) However, when explaining generalised neural networks, researchers tend to use dynamic scatter plots to display the working mechanism by showing the aggregation changes of the points. (see Figure 17)

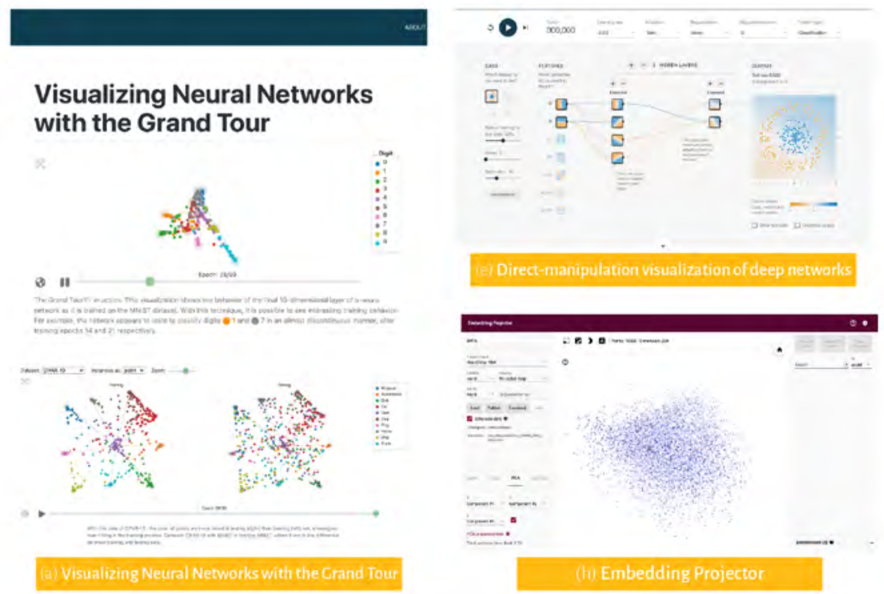


Figure. 17 Scatter plot

Conclusion

From the above analysis, we can see that almost all of the extant interpretation works on Machine Learning techniques come from computer scientists. They have used interactive visualisation of website, local visualisation toolkit and other forms to achieve insights sharing among peers with equivalent knowledge levels and knowledge dissemination with downward professional students. Given the communication objects have certain professional common sense, when the CS experts create the explanation work, mostly chooses the more scientific but the abstract form to present the professional content.

Plenty of researches are being discussed between professions, but little or nothing that concerns the requirement of the lay user. With the widespread application of Machine Learning in various industries with algorithmic errors that occur from time to time, it is increasingly indispensable for the vulnerable general public to understand how it works and to propose improvements as appropriate. Therefore, as communication designers, based on the premise that

users are ordinary people when designing interpretation of Machine Learning-related techniques, we will combine knowledge of information visualisation and narrative skills to allow audiences to initially understand concepts such as ConvNet by calling their shared expertise in life.

	Computer Scientist	**Communication Designer**
goal	Debug; Optimise the program; Share knowledge	Superficially demystify the decision-making process; Raise the awareness
target user	Professions; Students	Lay user
way of presenting	High degree of parameterisation	Information visualisation; Well-ordered narrative

Table. 3 Comparison of interpretation works from different standpoints

2.2 Learn from articles and videos online

In the initial investigation process, in addition to consulting journal articles, searching for related Popular Science articles and videos through Google was also a meaningful way to establish a knowledge structure about ConvNet. During searching, all the websites that have been viewed are entered into a sheet, which is divided into four categories according to the nature of the content:

- Wikipedia
- Explanation of ConvNet
- Applications of ConvNet
- Algorithmic errors related to ConvNet.

Wikipedia

Wikipedia provides the fastest way to get a definition of a term, and we can get to know a lot of authoritative papers from its reference section.

Explanation of ConvNet

The explanations of ConvNet obtained from the Internet in the form of but not limited to popular science articles, popular science videos and professional forum questions and answers. Content from the web is more approachable than the journal articles, whose creators use a large number of images, visualisations, diagrams and other media to help explain the object.

Applications of ConvNet & Algorithmic errors related to ConvNet

In view of the fact that the audience lacks professional knowledge, it is also crucial to investigate the application of ConvNet and related algorithm errors. Understanding the ConvNet applications in daily life gives users a more vivid sense of its importance while comprehending algorithmic errors make them fully aware of the potential seriousness of these errors and think about the relationship between the non-professions and the algorithms.

2.2.1 How do the online media explain ConvNet?

In this section, we are mainly discussing the online articles and videos explaining ConvNet. After conducting investigations on the Internet and making a note of browsing records. The following standards organised the condensed information in order to infer the potential audience: The purpose of the article (or video); The rhetoric of the article(script) and The type of media used to support the explanation.

Based on the degree of computer-science-literate, there are three levels:

1. Lay user;
2. Machine Learning or Computer Vision beginner;
3. Computer scientist.

In order to reduce the error caused by subjective judgment, the explanations were divided into five groups, as shown in table 4.1.

	Potential audience
Reading threshold = 1	1. Lay user
Reading threshold = 1.5	1. Lay user; 2. ML or CV beginner
Reading threshold = 2	2. ML or CV beginner
Reading threshold = 2.5	2. ML or CV beginner 3. Computer scientist
Reading threshold = 3	3. Computer scientist

Table. 4.1

1. Lay user
These are individuals who typically have no prior knowledge about ConvNet, and may not have a technical background. They simply use AI-powered devices and consumer applications.

2. ML or CV beginner
The audience who already have some experience and understand what ConvNet is. As neural network novices, they want to have in-depth study such as how to optimise a model.

3. Computer scientists
Their job is primarily focused on developing, experimenting with, and deploying deep neural networks. These model developers have a strong understanding of ConvNet techniques and a well-developed intuition surrounding the model building.

Information from the Internet

Title	What is useful?	# Tag	Aim of the article or video	Rhetoric	Supporting Media	Potential Audience	Link	Quotation	Classification
▼ Reading Threshold=1							6		
Image Classification With Convolutional Neural Network	the author works for Singapore national park, using cnn to classify birds species	#transer learning #VGGNet		-	- architecture - result showing accuracy - performance chart	1. Lay user	https://towardsdatascience.com/i code: https://github.com/JkYYnY		Reading Threshold=1
How Convolutional Neural Networks work	- Explaining without a lot of terms; - Mentioned some papers about CNN application	#video	Explain CNN	- enumerate - resume the computation	- diagram - visualisation	1. Lay user	https://www.youtube.com/watch?		Reading Threshold=1
Machine Learning for everyone (图解机器学习：人人都能懂的算法原理)	Explain these techniques in simple words.	#storyline #introduction	Explain machine learning technique to public	- enumerate	- diagram	1. Lay user	EN: https://vas3k. CN: https://mp.we	Machine Learn	Reading Threshold=1
How we teach computers to understand pictures | Fei Fei Li	Combined with rich media materials and introduced to the public in plain language "how to teach computer to understand the pictures"	#video #introduction #TED presentation	Explain CNN to public	- comparison (between little girl and computer: how they describe a image) - enumerate (using wired cats to present the different way that computer understand images)	- diagram - architecture - results showing accuracy	1. Lay user	https://www.youtube.com/watch?		Reading Threshold=1
一文看懂卷积神经网络	Contains 1. Unique value: 2. fundamentals: 3. practical applications and many good examples for presentation: CNN has been widely used, such as: face recognition, automatic driving, security and many other areas.	#beginner #introduction #storyline #application	explain CNN	- comparison (cnn vs human brain)	- diagram - visualisation - architecture - application	1. Lay user	https://easyai.tech/ai-definition/		Reading Threshold=1
通俗介绍卷积神经网络	long reference list	#introduction #layers	explain CNN: applications/ different layer's work/ some famous models	-	- results showing accuracy - architecture - diagram - visualisation	1. Lay user	https://www.pixb	Each picture is	Reading Threshold=1
▼ Reading Threshold=1.5							8		
Convolutional Neural Networks — A Beginner's Guide	using a series of diagrams and videos to explain the working process	#beginner #introduction #layers	This post is focused towards the final goal of implementing a MNIST handwritten digit classifier so everything is explained keeping that in mind — convolution layers, max pooling layers, ReLU activation function, fully connected layers, dropout layers, cross entropy loss function, etc.		- architecture - diagram - math formula	1. Lay user 2. ML or CV beginner	https://towardsdatascience.com/		Reading Threshold=1.5
Image Classifier - Cats 🐱 vs Dogs 🐶	nice narration. From human visual system to (R,G,B) to the operations of CNN.	#code #tutorial #GitHub repo #introduction	"How to classify images with Convolutional Neural Networks (CNNs) deployed on the Google Colaboratory platform. You will learn the essentials of the image classification systems and the underlying computer vision concepts."	- comparison (between Visual structure of the human brain and CNN)	- code snippet - architecture - diagram - code snippet - performance chart - visualisation	1. Lay user 2. ML or CV beginner	https://towardsdatascience.com/		Reading Threshold=1.5
Image Recognition: Dogs vs Cats! (92%)	Super detailed, the whole workflow could be understood.	#tutorial #GitHub repo #Jupyter notebook	Teach readers to build a CNN with high classification accuracy	-	- code snippet - classified images	1. Lay user 2. ML or CV beginner	https://thedatafrog.com/en/article		Reading Threshold=1.5
Beginner's Guide: Image Recognition and Deep Learning	-	#introduction	General explained "image recognition", "Deep learning" and "CNN"		- diagram	1. Lay user 2. ML or CV beginner	https://dzone.com/	A neural netwo For an AI appli	Reading Threshold=1.5
[CSDN] 深度学习卷积神经网络大事件一览	talked about the evolution of CNN	#timeline #important theories #introduction	-	-	- architecture	1. Lay user 2. ML or CV beginner	https://blog.csdn.net/u01308		Reading Threshold=1.5
[机器之心] 卷积神经网络	mentioned some applied areas	#introduction #history	General introduce CNN		- diagram - math formula	1. Lay user 2. ML or CV beginner	https://www.jiqizh	A convolutiona	Reading Threshold=1.5
卷积神经网络是什么? 你需要知道的三件事		#introduction #application #matlab	present by: definition/ working principle/ how to work with MatLab		- architecture	1. Lay user 2. ML or CV beginner	https://ww2.mathworks.cn/solutio		Reading Threshold=1.5
CNN笔记：通俗理解卷积神经网络	the way author explain "filter" is interesting! 1.[not good enough but works] For the information in the local area, human eyes are also biased and preferred. For example, looking at beautiful women, the focus is on the face, chest, and legs, so the weights of these three inputs are relatively large. 2.[super nice] When someone travels around the world, the information they see is changing, but the eyes collecting the information are unchanged. btw, the eyes of different people see the same local information differently, that is, a thousand readers have a thousand hamlets, so different filters are like different eyes, and different people have different feedback results.	#introduction #layers #neural network	Explanation from the overview of Neural Network to CNN		- architecture - math formula - diagram - visualisation	1. Lay user 2. ML or CV beginner	https://blog.csdn.net/v_JULY_v		Reading Threshold=1.5

Information from the Internet

Title	What is useful?	# Tag	Aim of the article or video	Rhetoric	Supporting Media	Potential Audience	Link	Quotation	Classification
▼ Reading Threshold=2							12		
A Beginner's Guide To Understanding Convolutional Neural Networks		#introduction	explain how CNN works	-	- architecture - diagram - performance chart - visualisation	2. ML or CV beginner	https://adeshpande3.github.io/A-I		Reading Threshold=2
Categories of algorithms 机器学习算法集锦：从贝叶斯到深度学习及各自优缺点	mind mapping a good way to sort out the complex knowledge system	#deeplearning #pros&cons #algorithm-overview	This article screens and briefly introduces some of the most common algorithm categories, and also lists some actual algorithms for each category and briefly introduces their advantages and disadvantages.	-	- architecture	2. ML or CV beginner	EN: https://static.coggle.it/diagra CN: https://mp.weixin.qq.com/s/h		Reading Threshold=2
Understanding Convolutions and Pooling in Neural Networks: a simple explanation	interesting examples about "filter"(kernel)	#convolution layer #pooling layer	explain Convolutions and Pooling in Neural Networks	-	- diagram - visualisation	2. ML or CV beginner	https://towardsda	A basic convol	Reading Threshold=2
How to Build a Dog Breed Classifier using CNN?		#dog breed #tutorial	how to train the Convolutional Neural Network (CNN) from scratch and get their accuracy to test the results on test data. The author also used different transfer learning techniques to improve model accuracy and get better classifications.		- code snippet - visualisation - architecture	2. ML or CV beginner	https://medium.com/swlh/how-to-		Reading Threshold=2
Having Fun Learning CNNs: Example of Dog Breed Prediction Applicable to Human Images	**interesting case!** when we feed in image of British Queen, it give the result "Maltese" :)	#dog breed	The application should provide a fun user experience as it will accept not only dog images but any user-supplied image as input. If a dog is detected, the algorithm should output a prediction of the dog's breed. If a human is detected, the output should be the most resembling dog breed!	-	- classified images - visualisation - code snippet	2. ML or CV beginner	https://towardsdatascience.com/		Reading Threshold=2
Intuitively Create CNN for Fashion Image Multi-class Classification		#fashion #tutorial	how to create a CNN for classifying different clothes step by step	-	- architecture - code snippet - classified images	2. ML or CV beginner	https://towardsdatascience.com/		Reading Threshold=2
一：卷积神经网络（CNN）在图像识别的最新进展		#introduction #models #image recognition	This article explores the latest progress of convolutional neural networks in image recognition	-	- architecture - visualisation - diagram	2. ML or CV beginner	https://zhuanlan.z	Image recognit	Reading Threshold=2
A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way	- Explanation about every layer	#introduction #layers	Explain CNN by analysing every layers	no	- architecture - diagram	2. ML or CV beginner	https://towardsdatascience.com/		Reading Threshold=2
ConvNetJS CIFAR-10 demo	A professor embedded a CNN into the browser.	#run algorithm #introduction	1. Inform the users how the images feature in every stage of operation. 2. Show the error rate of image recognition on a live running.	-	- diagram - architecture	2. ML or CV beginner	https://cs.stanford.edu/people/ka		Reading Threshold=2
Convolutional Neural Networks (CNNs) explained (with blogger)	"pattern" explanation is useful	#introduction #layers	explain how the convolution layers work	-	- diagram - architecture	2. ML or CV beginner	https://deeplizard.com/learn/vide		Reading Threshold=2
Image Classification of X-Ray Scans		#application #medical	introduce how to detect covid-19 by machine image classification	-	- architecture - performance chart	2. ML or CV beginner	https://towardsdatascience.com/		Reading Threshold=2
CNN卷积神经网络原理详解+图片识别应用（附源码）	super detailed tutorial for every step.	#introduction #layers #tutorial	explain how CNN works	- enumerate (cross stitch/ mouse image sample/)	- diagram - code snippet	2. ML or CV beginner	https://zhuanlan.z	The process of Using convolut Preliminary ext Main features c The fully conne Generate a cla	Reading Threshold=2
▼ Reading Threshold=2.5							7		
浅析图卷积神经网络	**intro is interesting!!: could be useful while presentation** If you have an image, to do the classification, the traditional method needs to extract some features manually, such as texture, color, or some more advanced features. These features are then placed into a classifier such as a random forest and given an output label that tells it which category it is. Deep Learning, on the other hand, inputs a graph, passes through a neural network, and outputs a label directly. Feature Extraction and classification in one step, to avoid manual extraction of features or manual rules, from the original data to automatically extract features, is an end-to-end learning. Compared with traditional methods, deep learning can learn more effective features and patterns.	#GCN(graph convolutional network) #Euclidean data	explain graph convolutional network	-	- diagram - math formula - performance chart	2. ML or CV beginner 3. Computer scientists	https://mp.weixin.	Convolutional	Reading Threshold=2.5
CNN Architectures : VGG, ResNet, Inception + TL	The way to test the performance of different architectures in parallel is interesting	#tutorial #imageNet #transfer learning	Compare popular CNN architectures; Explained how to generate image features using pre-trained models	- comparison (same image set trained by different architectures to see different performances)	- architecture - diagram - code snippet	2. ML or CV beginner 3. Computer scientists	https://www.kaggle.com/shivamb		Reading Threshold=2.5
[知乎] CNN简史	talked about the evolution of CNN	#timeline #introduction #important theories	sort out the evolution of CNN development	-	- architecture	2. ML or CV beginner 3. Computer scientists	https://zhuanlan.zhihu.com/p/390		Reading Threshold=2.5
A Recipe for Training Neural Networks 33 个神经网络「炼丹」技巧	The post tells us how to improve the quality of neural networks from a special perspective	#improvment		- metaphor (DL process -> cooking elixirs)	- code snippet	2. ML or CV beginner 3. Computer scientists	EN: https://karpathy.github.io/201 CN: https://zhuanlan.zhihu.com/p		Reading Threshold=2.5
Keras Vs PyTorch Dilemma Keras vs PyTorch，哪一个更适合做深度学习?	This article compares two softwares: Keras and PyTorch	#tool #frameworks	Compare two frameworks by training a same model	-	- code snippet - performance chart	2. ML or CV beginner 3. Computer scientists	EN: https://medium.com/@karan_ CN: https://mp.weixin.qq.com/s/F		Reading Threshold=2.5
Review of LeNet-5: How to design the architecture of CNN	Discuss how to design cnn based on known classic model (LeNet)	#LeNet	Incorporate knowledge to tailor the architecture of CNN	-	- architecture - visualisation	2. ML or CV beginner 3. Computer scientists	https://towardsdatascience.com/		Reading Threshold=2.5
卷积神经网络 CNN BP算法推导		#backpropagation	present the process of back propagation's derivation	-	- architecture - diagram - math formula	2. ML or CV beginner 3. Computer scientists	https://www.cnblogs.com/chenjie		Reading Threshold=2.5

Information from the Internet

Title	What is useful?	# Tag	Aim of the article or video	Rhetoric	Supporting Media	Potential Audience	Link	Quotation	Classification
▼ Reading Threshold=3							9		
Convolutional Neural Networks (LeNet)		#not for beginner	Explain CNN by analysing LeNet deeply	-	- architecture - code snippet - visualisation	3. Computer scientists	http://deeplearnin	Simple cells re The animal visu	Reading Threshold=3
Face Recognition using Deep learning		#face recognition #Attendance Record	Author would like to develop a face recognition system that will be used within a class as an attendance system to mark presence of lecturers and students.		- architecture - visualisation - math formula	3. Computer scientists	https://towardsdatascience.com/		Reading Threshold=3
THE HISTORY OF NEURAL NETWORKS	the first diagram is useful to present an overview of different models	#history #evolution #introuction	This article compares various well-known models by exploring the neural network architecture and computing performance	-	- architecture - diagram - performance chart	3. Computer scientists	http://dataconomy.com/2017/04/		Reading Threshold=3
Geoffrey Hinton: What is wrong with convolutional neural nets? | Fields Institute, 2017		#imperfect #improvment	Professional speech about the CNN, how to improve cnn by optimising its structure	-	-	3. Computer scientists	https://youtu.be/Mqt8fs6ZbHk Distilled note: https://medium.com		Reading Threshold=3
Loss and Loss Functions for Training Deep Learning Neural Networks 理解计算机视觉中的损失函数	In this post, you discovered the role of loss and loss functions in training deep learning neural networks and how to choose the right loss function for your predictive modeling problems.	#Loss function #optimization	Neural networks are trained using an optimization process that requires a loss function to calculate the model error. Maximum Likelihood provides a framework for choosing a loss function when training neural networks and machine learning models in general. Cross-entropy and mean squared error are the two main types of loss functions to use when training neural network models.	-	- code snippet	3. Computer scientists	EN: https://machinelearningmaste CN: https://mp.weixin.qq.com/s/E		Reading Threshold=3
CNN结构设计技巧-兼顾速度精度与工程实现	I can't understand.	#optimisation	This blog post discusses how to optimize the structure of cnn from different application perspectives	-	- architecture - diagram - visualisation	3. Computer scientists	https://mp.weixin.qq.com/s/loHW		Reading Threshold=3
Image Segmentation: Tips and Tricks from 39 Kaggle Competitions 在参加了39场Kaggle比赛之后，有人总结了一份图像分割炼丹的「奇技淫巧」	The gif on top is interesting!!	#image segementation #optimisation	-	-	-	3. Computer scientists	EN: https://neptune.ai/blog/image CN: https://mp.weixin.qq.com/s/n		Reading Threshold=3
【CNN结构设计】无痛的涨点技巧：ACNet	ACNet: structural improvement	#optimisation	Introduce ACNet: which could enhance the robustness of the model to image rollover and rotation	-	- diagram - architecrure	3. Computer scientists	https://mp.weixin.qq.com/s/ZKjH involved paper: https://arxiv.org/p		Reading Threshold=3
【CNN结构设计】深入理解深度可分离卷积			introduce a new CNN design	- comparison (standard cnn vs depthwise + pointwise cnn)	- architecture - code snippet	3. Computer scientists	https://mp.weixin.qq.com/s/IZ-nb		Reading Threshold=3

Table. 4.2 Record of the website explaining ConvNet

Furthermore, after condensing the information, “Potential Audience” and “Supporting Media” were extracted for making visualisation, as shown in figure 18. Since the target audience for the subsequent design work is the lay user, the “Supporting Media” associated with the Lay user is highlighted here.

Regarding the upper cluster “1. Lay user”, compared with standard academic architectures, the ConvNet interpretation based on web blogs, uses more common diagrams to reveal how this neural network works. Very few articles(or videos) also use statistical formulas, illustrations and screenshots of the test charts when running the program.

As for the lower cluster “1. Lay user 2. ML or CV beginner”, these articles(and videos) use a similar number of academic architectures and common diagrams explaining the convolution, pooling, fully-connected process. Similar to the cluster above, the blogs in this group also use the scientific formulas and test screenshots.

The most significant difference between the two clusters is that the second cluster uses code snippets for the explanation because its target audience includes ML beginners who actually need to run the neural network by creating code.

2.2.2 The math behind

In the previous chapter(2.1.2), we enumerated several classic ConvNets models. It is clear to find that the visualisations of ConvNets from the papers of Computer Scientists are layered structures. However, this kind of neural networks are just different code snippets that strictly follow the order; the output of one becomes the input of the next. The researchers use the vivid concept of the layer to symbolise the operation in ConvNet, a multi-layer neural network in which convolutional layers are convolution operations; pooling layers are pooling operations; fully-connected layers are fully connected operations.

Simply put, the convolution operation is mainly used to extract the various features of the Euclidean Data[16], followed by the pooling operation, which expands the main features to weaken the subtle features. Finally, the fully-connected operation is used to integrate those features.

[16] The most prominent feature of Euclidean data is the regular spatial structure and grid-like topology. For example, the picture is a regular square grid, and the speech is a regular one-dimensional sequence. These data structures can be represented by one-dimensional and two-dimensional matrices, and convolutional neural networks are very efficient in processing.

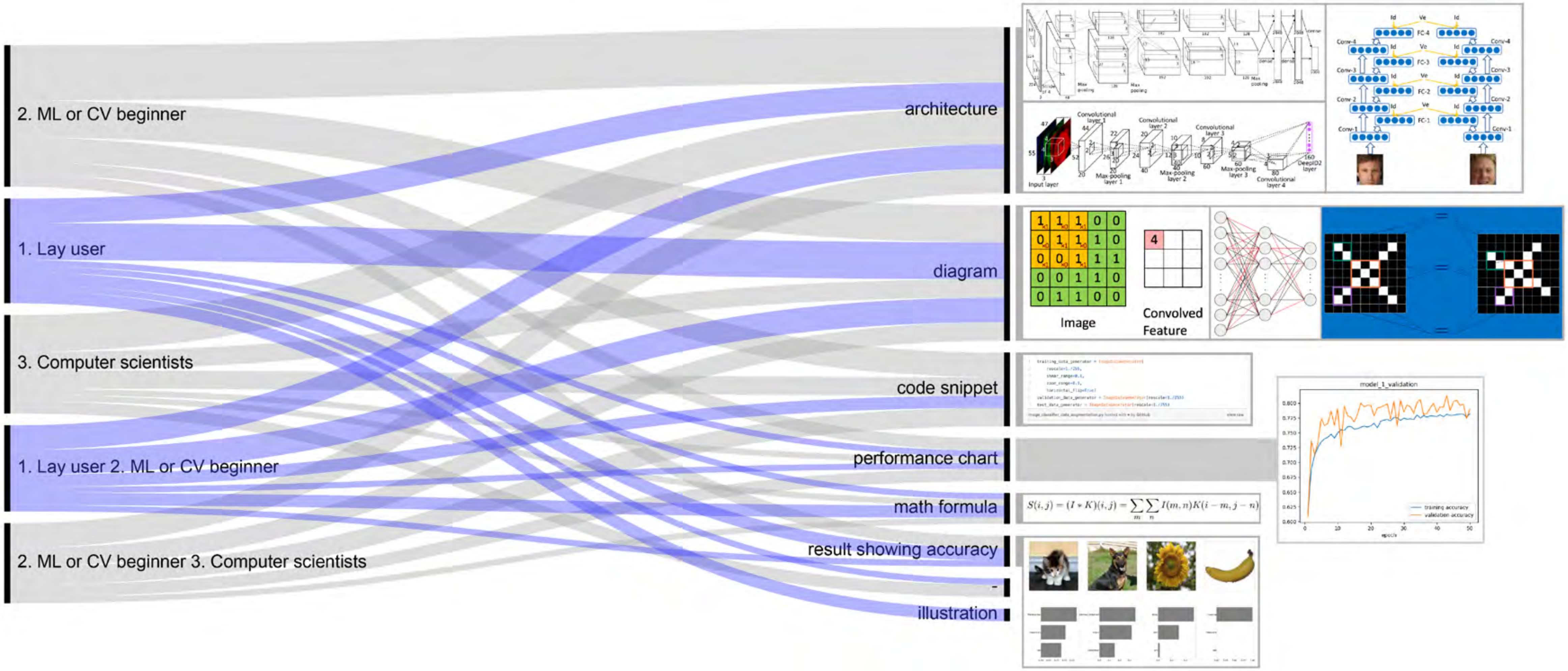
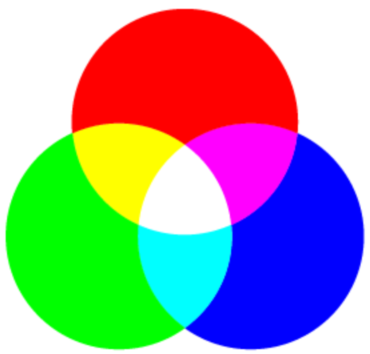


Figure. 18 Types of media used while explaining ConvNet to the lay users

The human eye is sensitive to the three colours of RGB(Red, Green, Blue), so the display also displays in these three colours, which makes most of the digital images describe colours in the three channels. Images are stored as pixels inside the computer. Each image is made up of countless pixels, each of which is assigned a value that can be digitally quantised by RGB.



	(Red, Green, Blue)
Red	(255, 0 , 0)
Yellow	(255,255, 0)
Green	(0 ,255, 0)
Cyan	(0 ,255,255)
Blue	(0 , 0 ,255)
Magenta	(255, 0 ,255)
White	(255,255,255)
Black	(0 , 0 , 0)

Figure. 19 RGB colour model

Humans recognise objects with their eyes, but for machines, an image is just a two-dimensional Pixel Matrix.

If we want to teach a computer to recognise cats, we will use mathematical language to let the computer understand what a cat is. Recognise the outline of the cat in the image above? No, the computer would not recognise the cats as shown in Figure. 21.

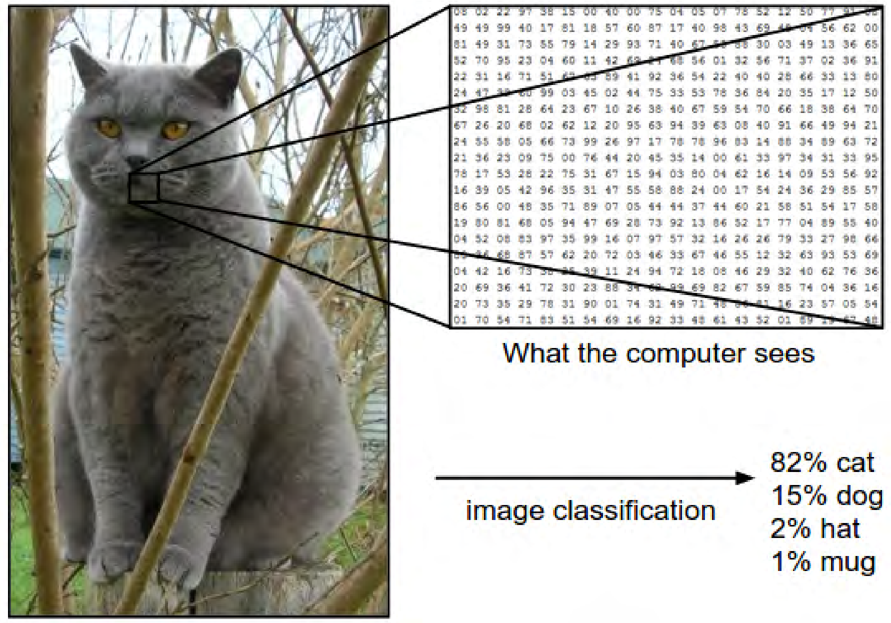


Figure. 20 Source: https://cs231n.github.io/assets/classify.png

Figure. 21 Source: https://www.pexels.com/photo/view-of-cat-in-snow-248276/ https://pixabay.com/p-393294/?no_redirect https://commons.wikimedia.org/wiki/File:New_hiding_place_(4224719255).jpg

As feature extractors, ConvNets are unique in extracting different features and then matching them locally according to a specific vote ratio. Therefore, we should let the computer learn the characteristics of cats, such as almond-shaped eyes, long tail, flat round head and triangular ears. Only when several of the above characteristics are met at the same time can it be considered as a cat.

In the following simple example, we will introduce the three operations (Convolution/ Max-pooling/ Fully-connected) in detail.

Checkerboard deduction[17]

We are training the easiest ConvNet for recognising, whether it is O in this image. This 9x9 checkerboard is a two-dimensional array as a bunch of ones and minus ones.

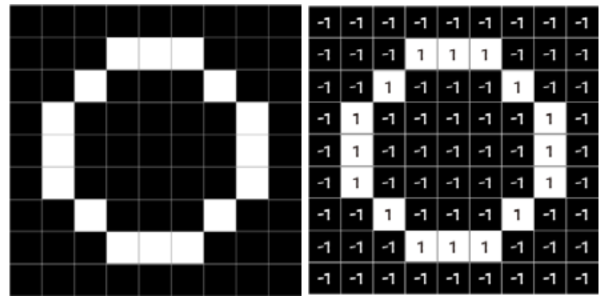


Figure. 22 Figure. 21 standard sample of letter "O" in 9*9 checkerboard

Firstly, we extract 4 **Feature**s of O from the standard image, which are used to locate a part of O. The **Feature**s are also known as the convolution **Kernel**s on ConvNets.

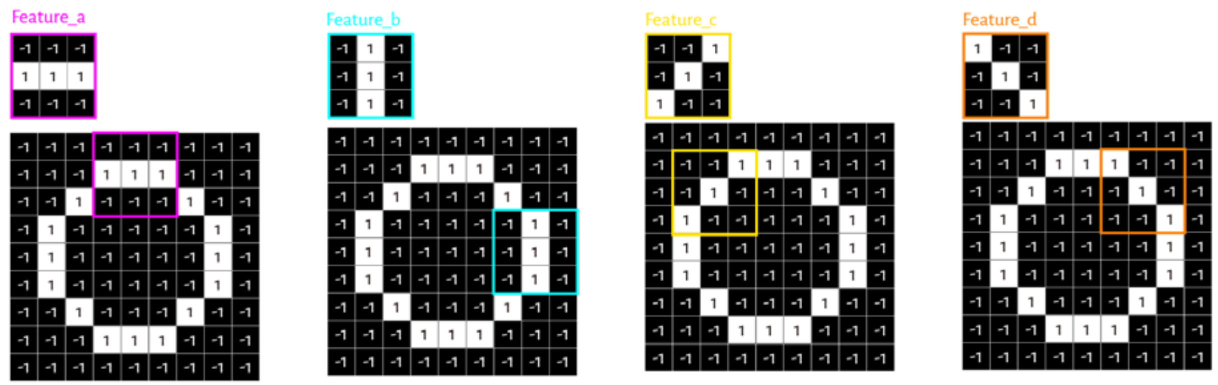


Figure. 23 four features of "O" were extracted manually

[17] This section inspired by How Convolutional Neural Networks work & How convolutional neural networks work, in depth

- *Convolution Operation*

After feature extraction, it is time to perform the convolution operation, which is simply the corresponding multiplication. As shown below, take the (1,1) element value in the green box from the **Feature_a**, and then take the (1,1) element value in the blue box from the standard image, and the two are multiplied to equal 1. Fill the result 1 into the new figure. After repeating this nine times, we get nine values.

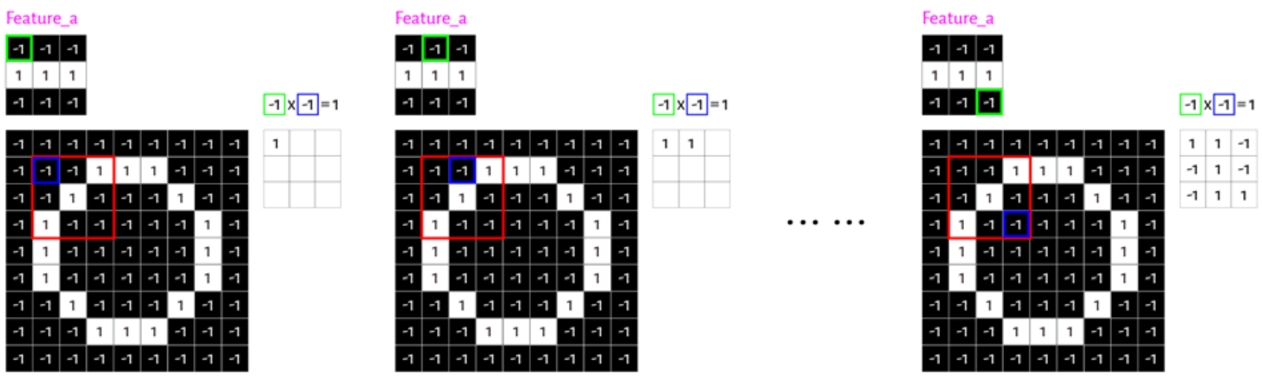


Figure. 24.1 Convolution computation(1/2)

Then average these nine values to get a new value, and fill the new value into another 7x7 graph. This new map is called the Feature Map. After the convolution corresponding multiplication and average, the red sliding window begins to slide to the right. The sliding range is selected according to the different Stride. For example, if stride=1, it will shift one pixel at a time. As shown in the figure below, the stride of the convolution operation for Feature_a is 1.

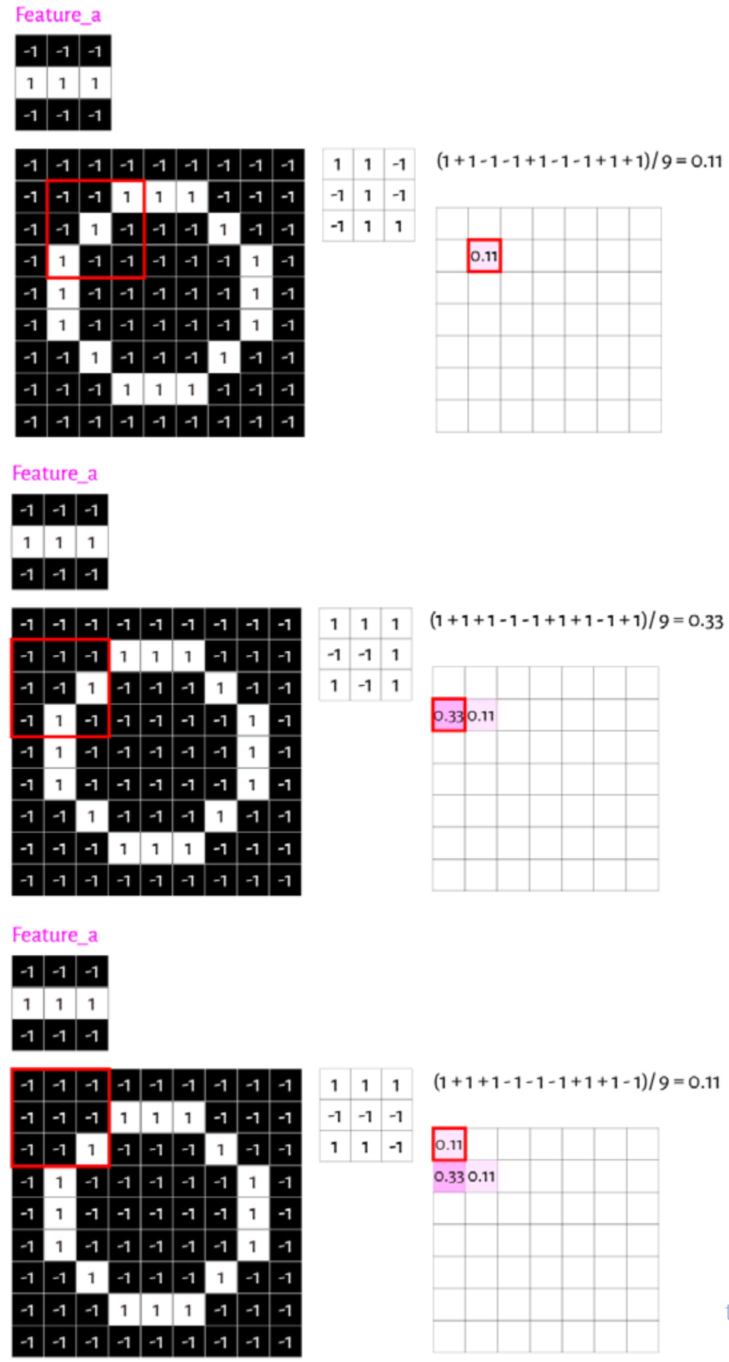


Figure. 24.2 Convolution computation(2/2)

By sliding the red boxes in the standard image, we run the calculations over and over until we fill a **Feature Map**[18] filtered by **Feature_a**.

[18] A Feature Map extracted from the image by a specific feature. The closer the value is to 1, the more complete the match between the corresponding position and the feature; the closer it is to -1, the more complete the match between the corresponding position and the opposite side of the feature, a value close to 0 indicates that there is no match or correlation.

As shown in the figure below, **Feature_a** is a horizontal-line kernel. The top and bottom colours in the **Feature Map** are darker (that is, the values are closer to 1), indicating a complete match between the top and bottom positions of the image and **Feature_a**.

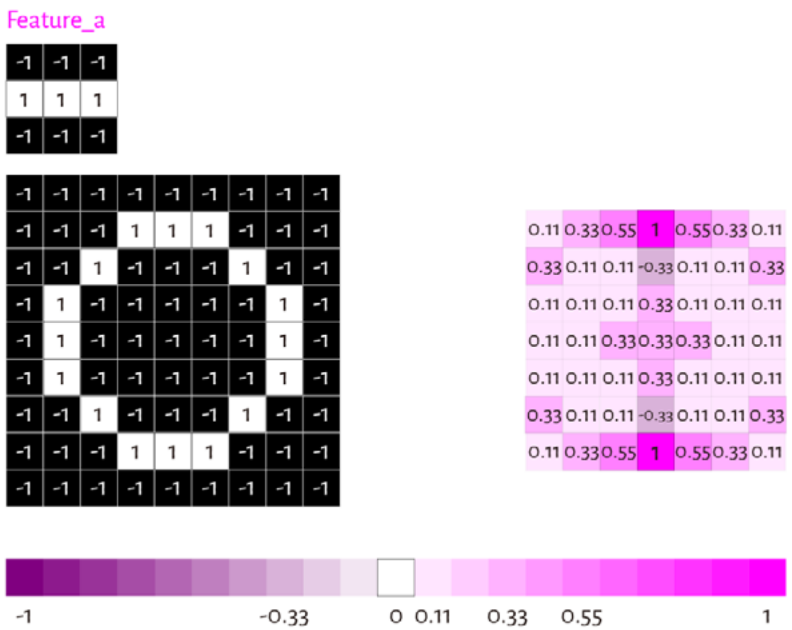


Figure. 25 Feature Map of "O" extracted by the horizontal-line kernel

A typical feature acts on the image to generate a feature map. For the O image, we have extracted four features, so four Feature Maps are finally generated.

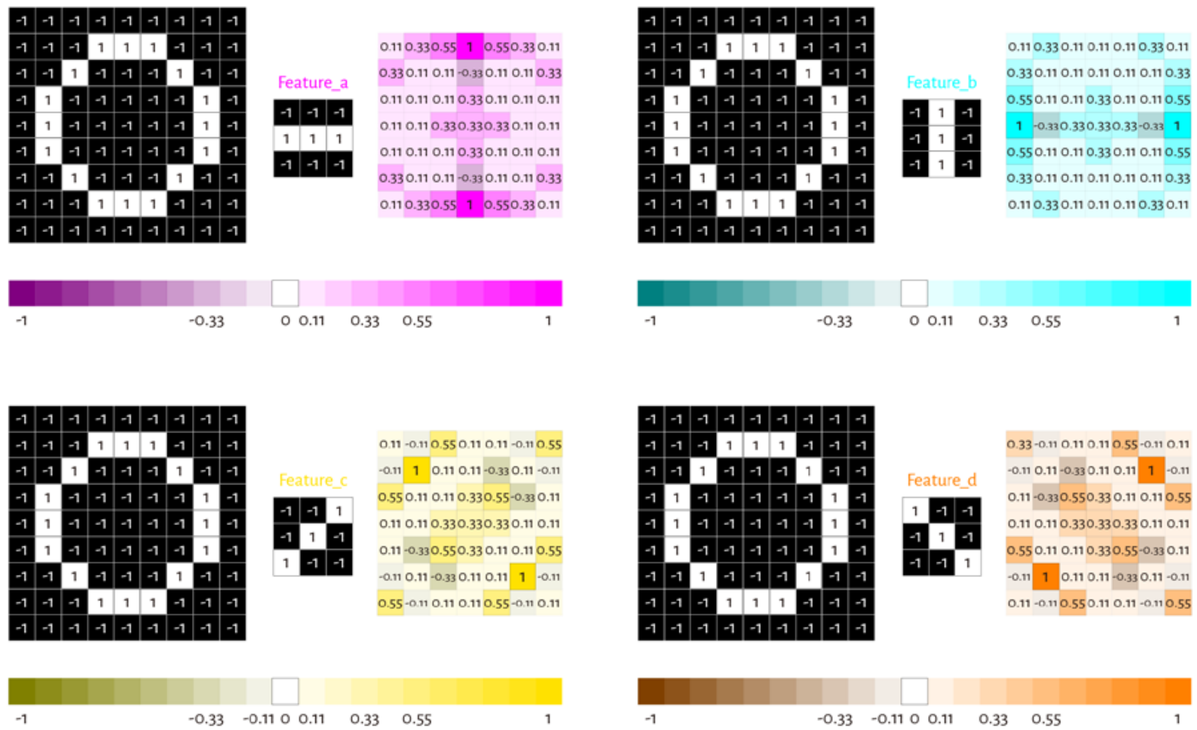


Figure. 26 four Feature Maps after convolution operation

- *Pooling Operation*

Figure. 27 Max-Pooling operation

After the convolution operation, we get Feature Maps with different values. Although the amount of data is much less than the original image, it is still too large. (compared to deep learning, there are hundreds of thousands of training images) Therefore, the following pooling operation can play a role. Its most prominent goal is to reduce the amount of data.

There are two types of pooling, Max-Pooling and Average-Pooling. As the name implies, Max-Pooling is to take the maximum value, and Average-Pooling is to take the average value. Take the example of Max-Pooling: choose the pooling size like 2x2, (that is, a 2x2 window is selected), and the maximum value is selected and updated into the new Feature Map, as shown in the figure above.

These four pooled Feature Maps show a significant reduction in the amount of data. (Figure.28) After performing Max-Pooling operation again, we get four 2 * 2 Matrices. (Figure.29)

Max-Pooling operation preserves the maximum value within each small block, so it is equivalent to retaining the best matching result of the block. Instead of focusing on exactly what matches in the window, it focuses on whether there is a match.

It could also be seen that ConvNets are able to find out whether there is a particular feature in the image, regardless of where it is present.

Figure. 28 Feature Maps changed after Max-Pooling operation

Figure. 29 Max-Pooling operation again

- *Fully-connected Operation*

Fully-connected layer, as the name suggests, is all connected. The convolution operation has the "local connection" idea, which uses a 3x3 image to connect to the original image. It is clear that there is only one 3x3 window in the original image that can be connected to it.

The unconnected part outside the window is connected later by sliding the window up. The idea of this method is “parameter sharing”; the parameter refers to the Kernel. They are using a sliding window to share the filter value with each region of the original image for convolution. Local connections and parameter sharing are two of the essential properties of ConvNets.

The Fully-connected operation is to summarise all the previous operations and present a final result. Its most significant purpose is to change the dimension of the Feature Map to get the vote corresponding to each classification category.

Now, every value gets a vote on what the answer is going to be. We break the 2x2 Feature Maps out, rearrange and put them into a single list. Furthermore, each of those connects to one of our answers that we are going to vote for. When we feed this in O, there will be specific values that tend to be high and predict very strongly. They get many votes for the O outcome.

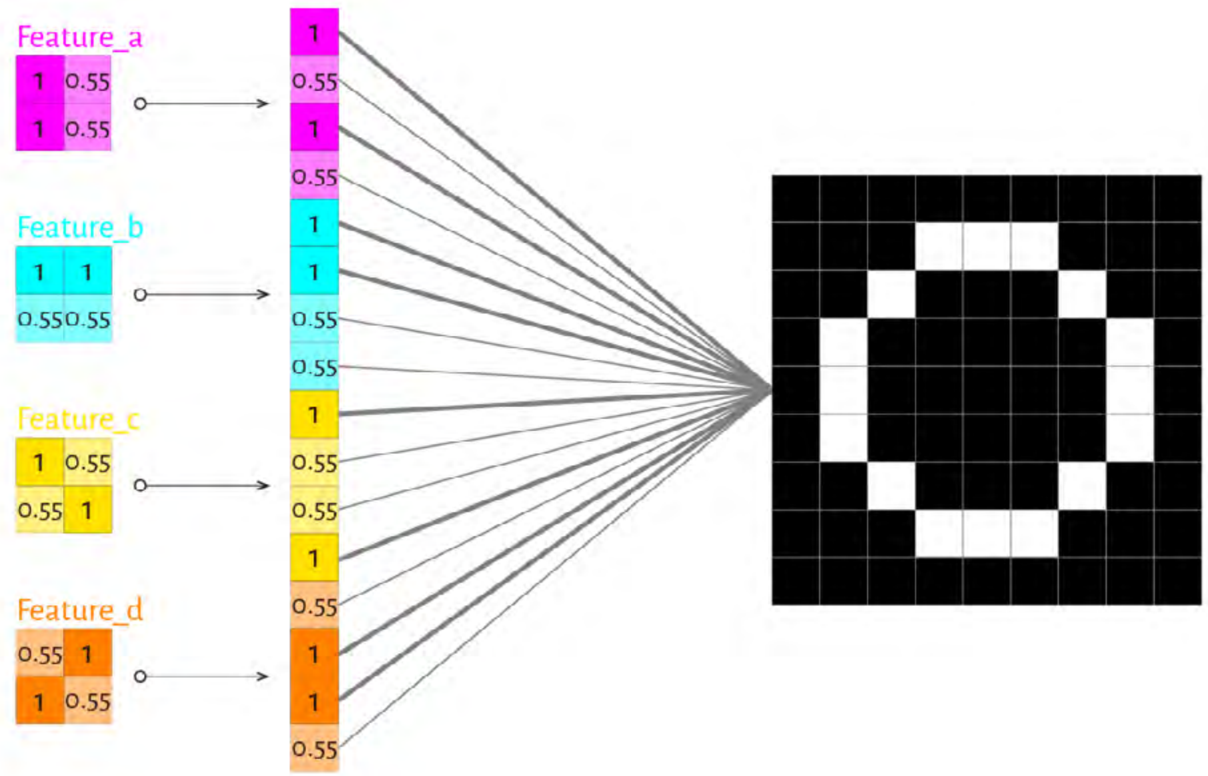


Figure. 30 A batch of votes indicating the features of “O”, in which the 1st, 3rd, 5th, 6th, 9th, 12th, 14th, 15th values equal to 1 and strongly predict that this letter is “O”.

- *Summary*

Let us revisit the process from the beginning. First, we analysed the training image, extracted the local features from it, and used these features as filters to perform convolution operations on the training image. After obtaining the Feature Maps, performed two times Max-Pooling operations to shrink the data volume. Finally, the fully-connected operation was used to calculate the weight of each feature vote macroscopically.

The below steps are Training the neural networks. Now we can say that the computer has more or less figured out the way to recognise “O”: A random alphabetic image that goes through the four operations shown above and outputs a sequence of numbers. Based on the result above, we could proclaim that the closer the average of individual numbers in the sequence is to 1, the more likely it is that the letter is “O”.

Now we get new inputs (“V”, “Q” and rotated“O”) and we want the simply-built ConvNets to identify whether they are “O” or not. In this way, these three inputs go through all of our 4-layer ConvNets, and then we get a series of votes. Based on the weights that each value gets to vote with we get average votes at the end. (as shown in the upper right corner of Figure. 32.1~32.3)

Training Process

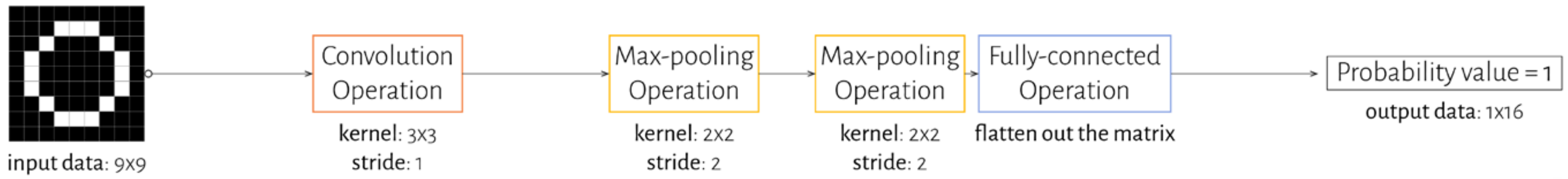
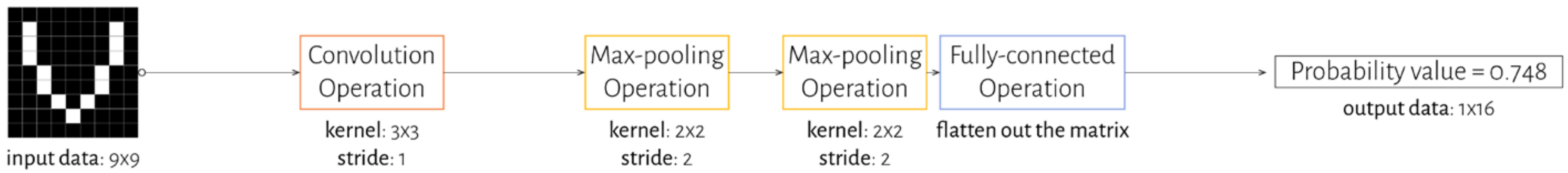


Figure.31 Training ConvNets to recognise "O"

Validating Process



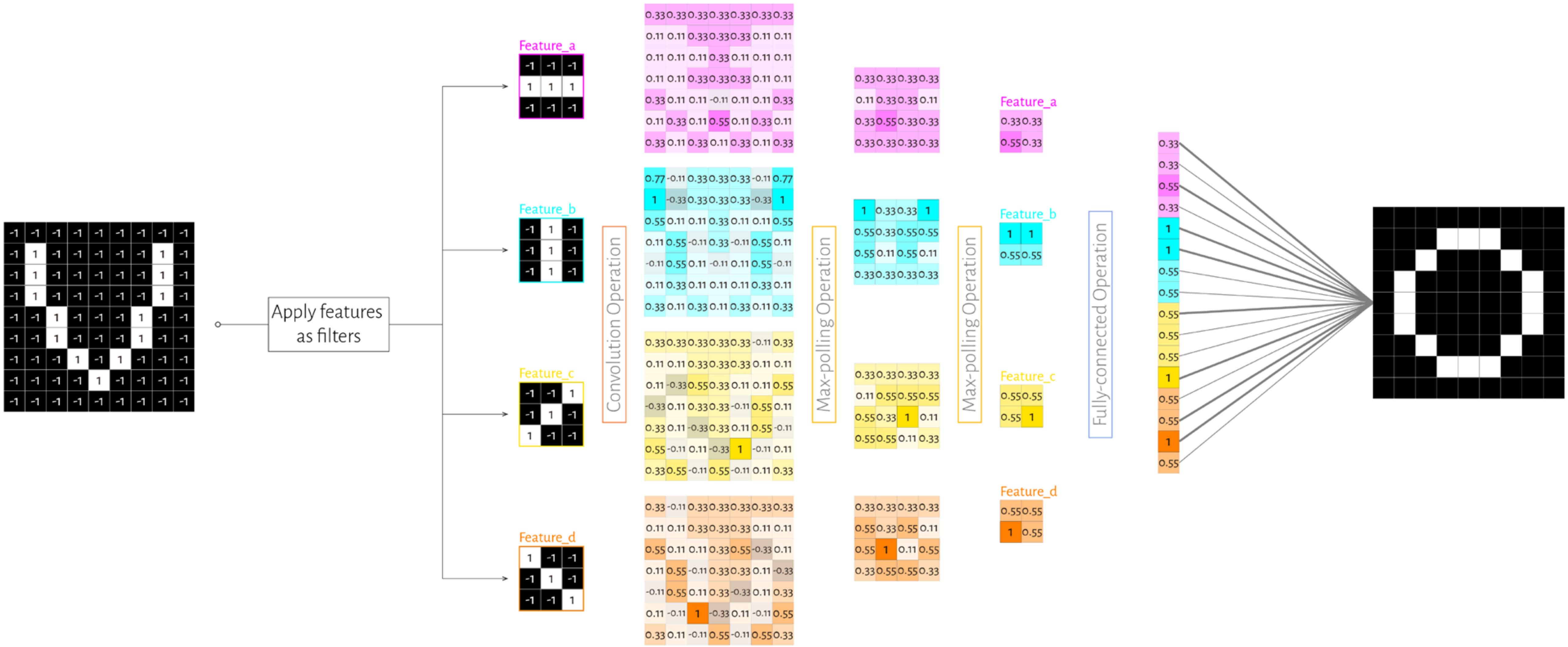
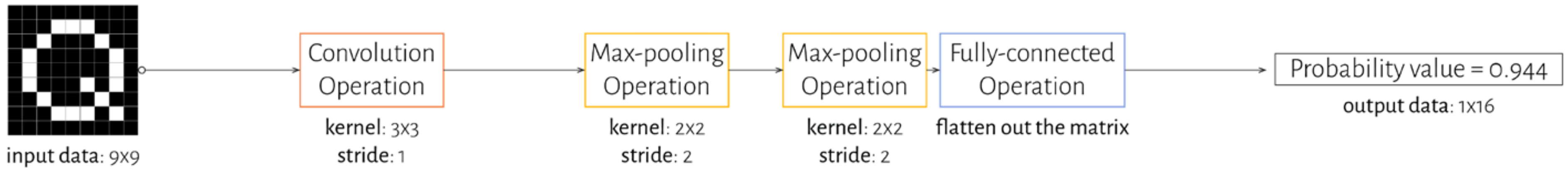


Figure. 32.1 Validating a random letter (1/3)

Validating Process



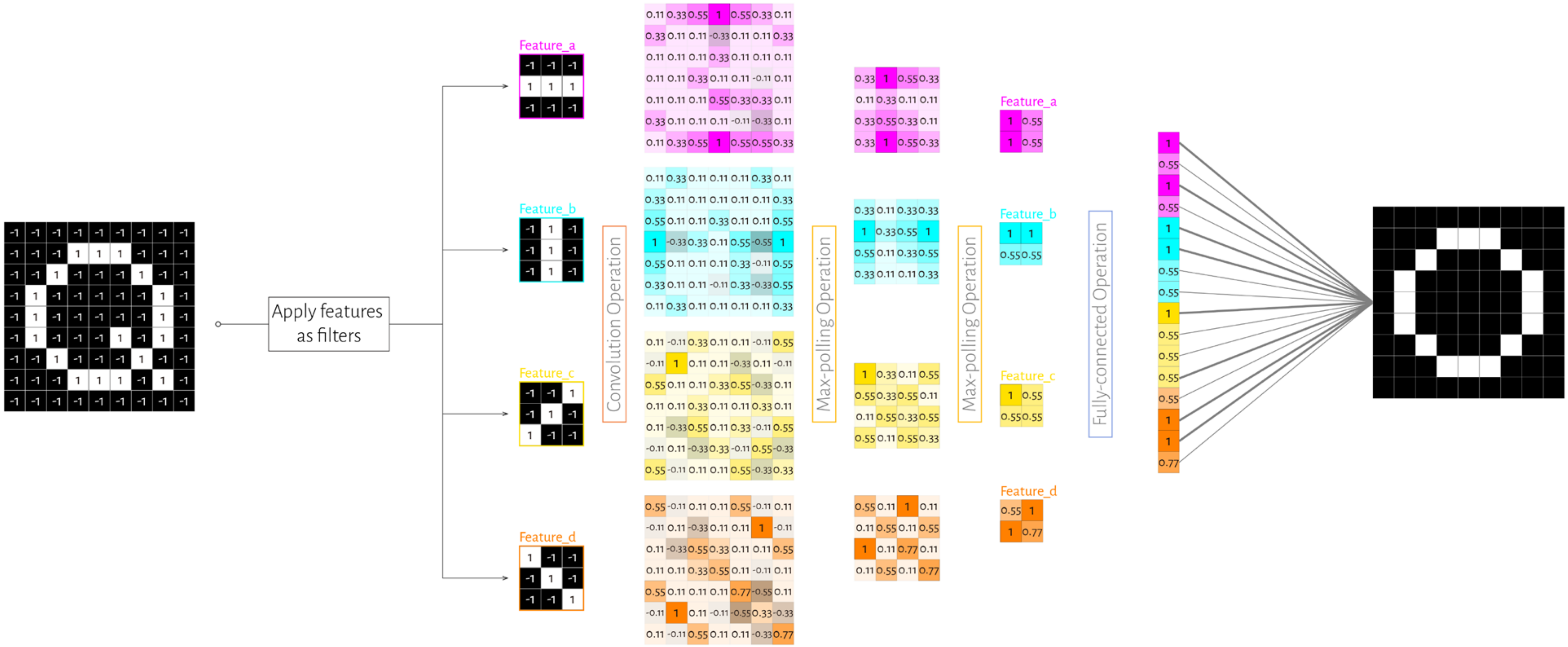
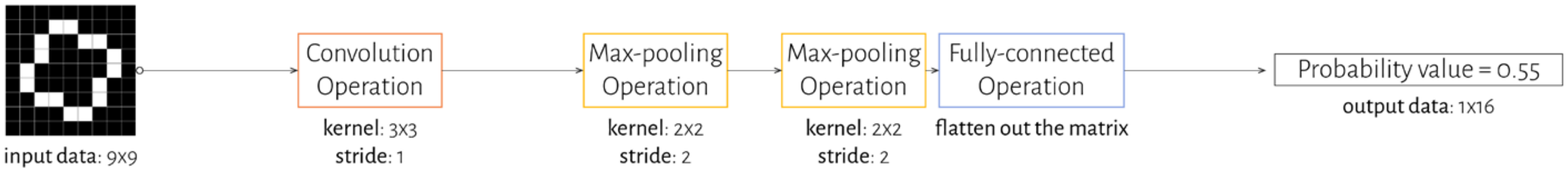


Figure. 32.2 Validating a random letter (2/3)

Validating Process



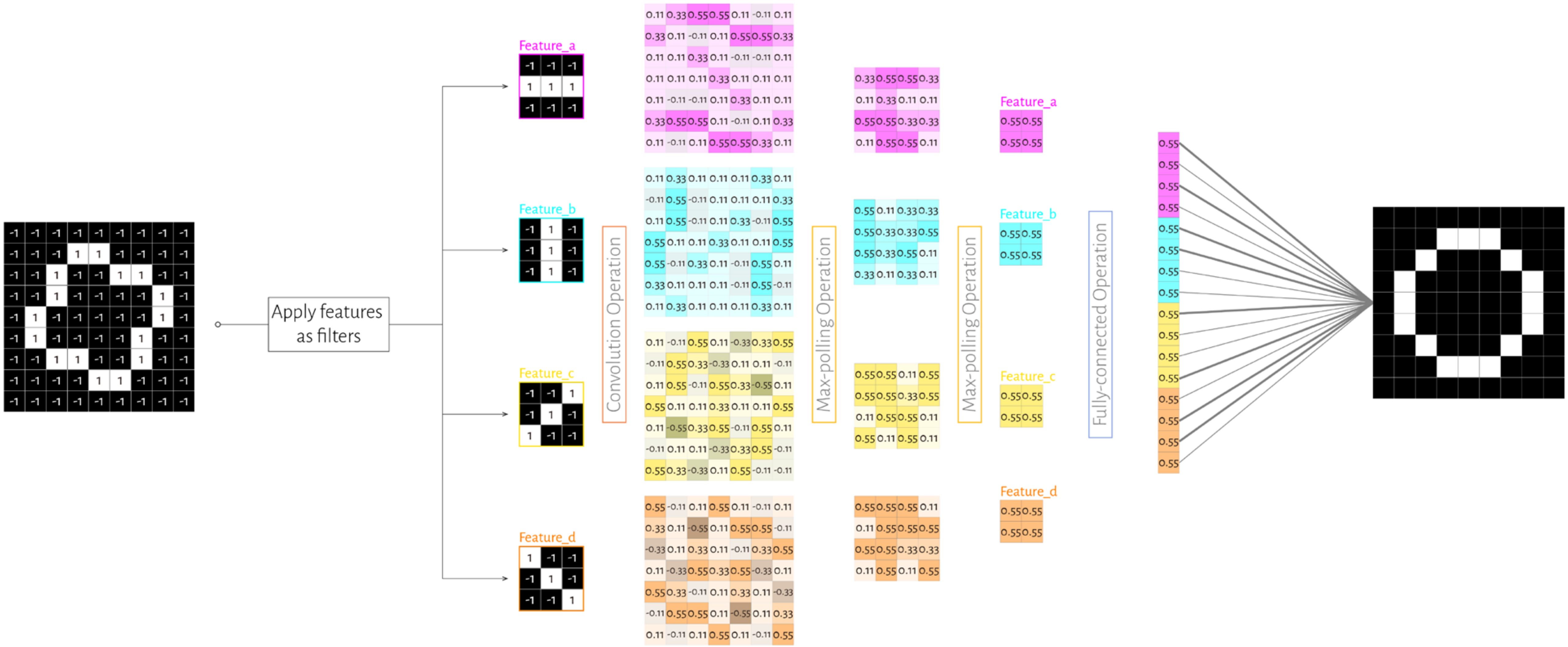


Figure. 32.3 Validating a random letter (3/3)

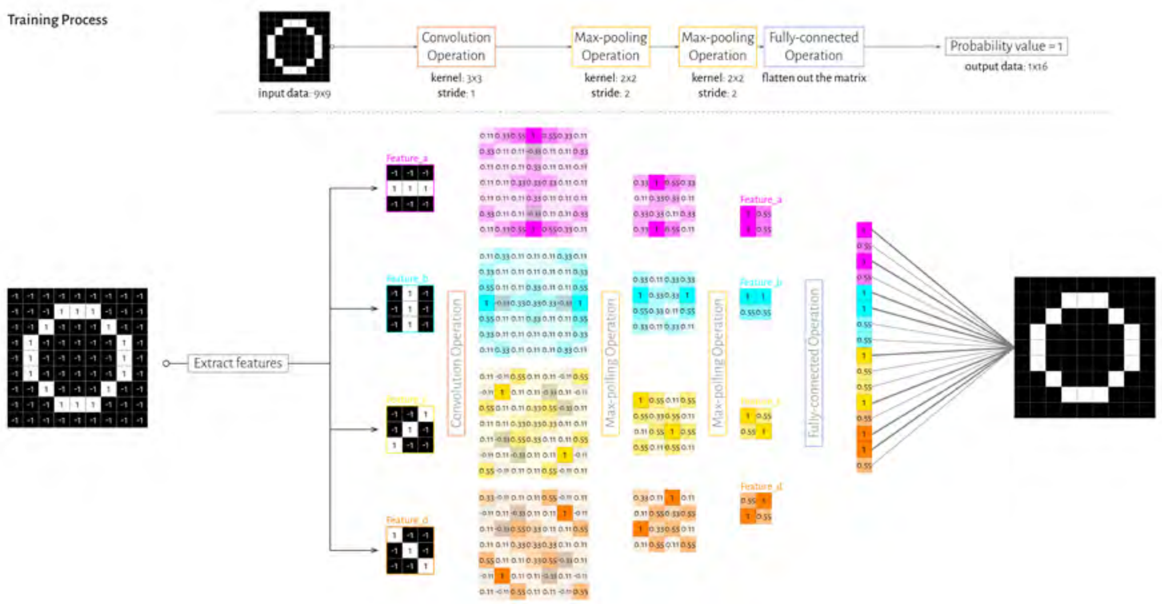


Figure.31 Training ConvNets to recognise "O"

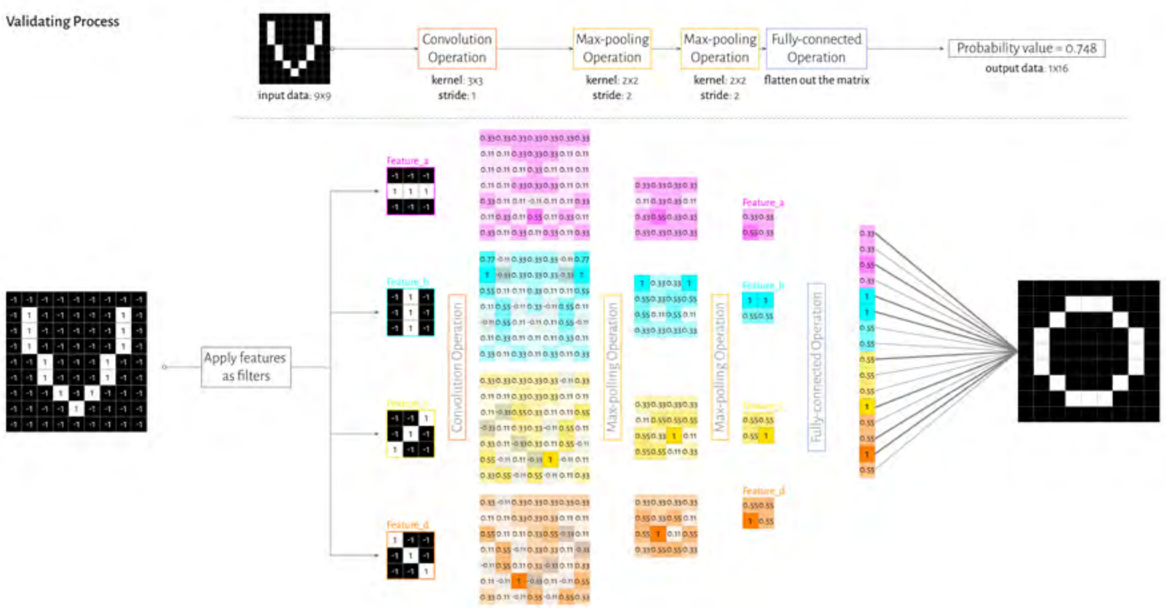


Figure. 32.1 Validating a random letter (1/3)

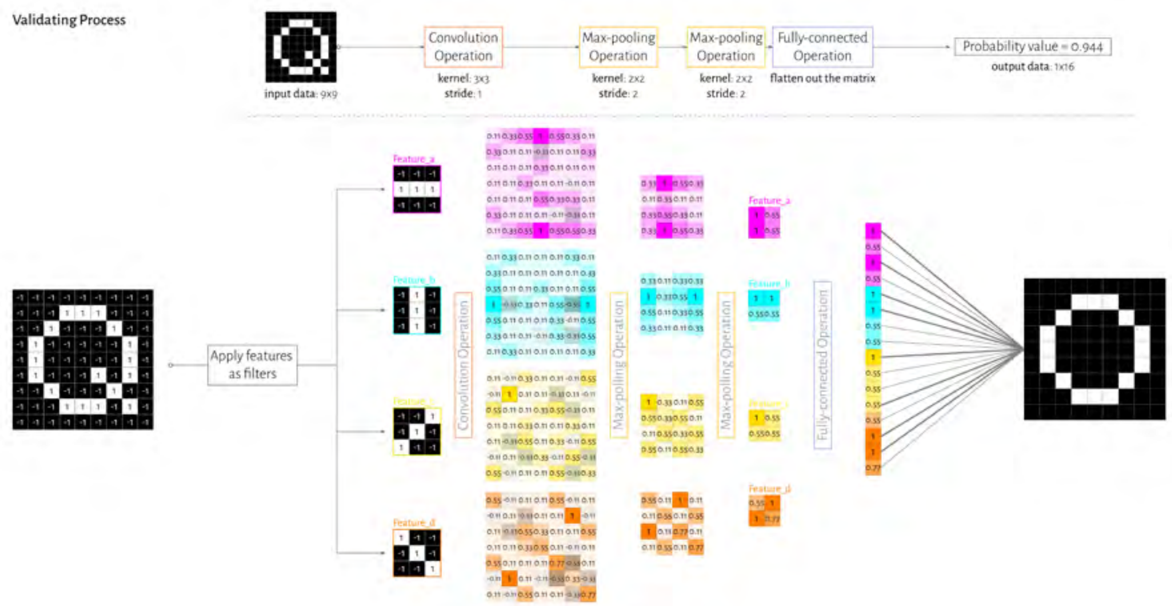


Figure. 32.2 Validating a random letter (2/3)

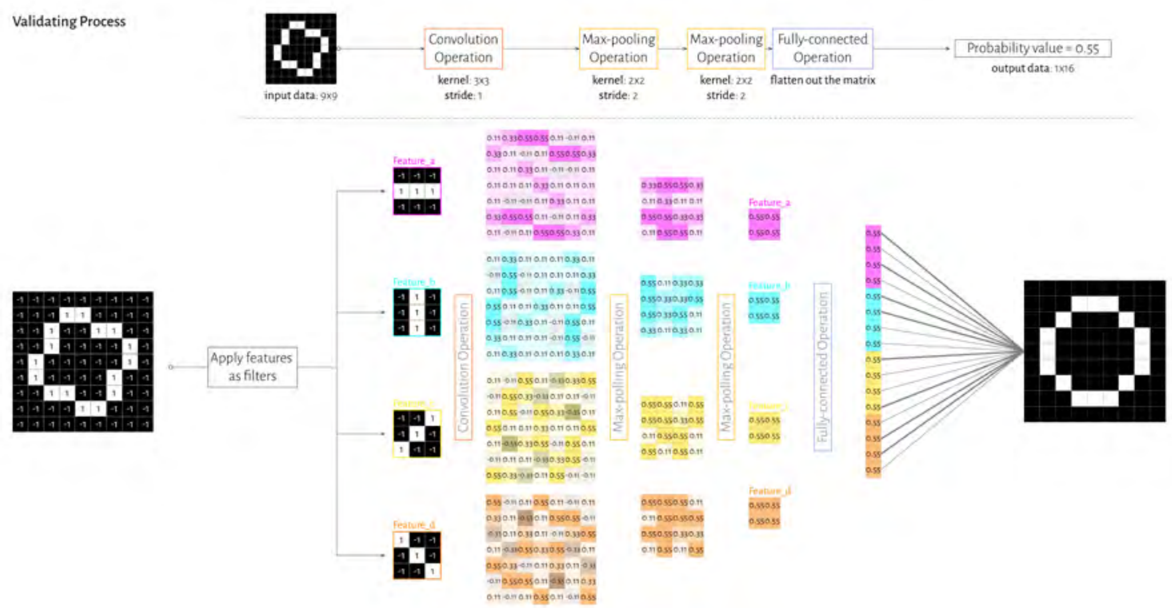


Figure. 32.3 Validating a random letter (3/3)

After summing up all the probability values and the input pictures, some problems appeared.

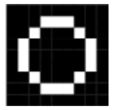
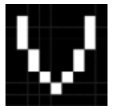
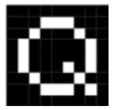
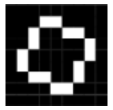
	Probability of being recognised as "O"
	1 (standard value)
	0.748
	0.944
	0.55

Table. 5 Deduction result summarisation

In previous sections, we have discussed that ConvNets recognise by matching local details, so in theory, actions such as scaling and rotation will not affect the recognition results. However, the fact is that the probability value of the rotated letter O is only 0.55, even lower than that of the letter V (0.748). On the other hand, since the upper part of the letter, Q is exactly the same as the sample O, the probability value is as high as 0.944.

The finding reveals a problem with neural networks in this deduction process: Too few training samples lead to insufficient feature extraction.

After all, the above is just a trivial experiment, which aims to demystify three types of computation of ConvNets. The networks created do not even deserve to be called ConvNets. When actually applying ConvNets, engineers do not need to analyse the features themselves. Thanks to the Back-Propagation[19] (Rumelhart et al., 1986) algorithm and massive training datasets, ConvNets uses a large number of iterations to modify the weights to reduce the error rate continuously.

So the underlying principle of Back-Propagation is the error in the final answer, which is used to determine how much the network adjusts. For each of these magic values, each of the feature pixel and each voting weight, they are affected by Back-Propagation and adjusted up and down to see how the error changes. The amount that they are adjusted is determined by how big the error is: the large error they are adjusted a lot, the small error just a tiny bit, no error they are not adjusted at all.

[19] When we use a feedforward neural network to accept an input x and produce an output $\hat{y}$, information flows forward through the network. The inputs x provide the initial information that then propagates up to the hidden units at each layer and finally produces $\hat{y}$. This is called forward propagation. Back-Propagation, often simply called backprop, allows the information from the cost to then flow backwards through the network, in order to compute the gradient. (Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. , 2016, p.202)

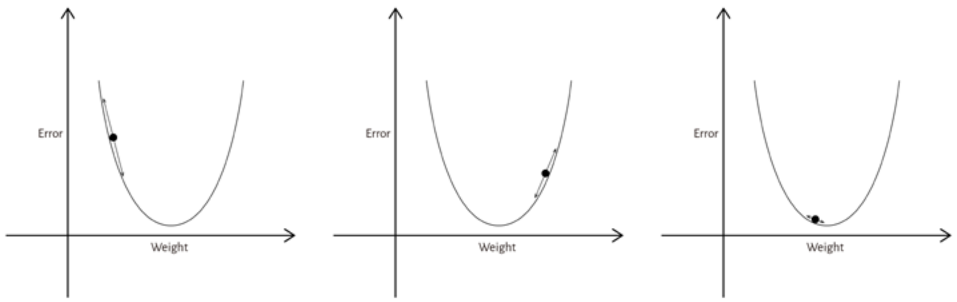


Figure. 33 relationship between Error and Weight in Back-Propagation

As they are adjusted, we can think of that as sliding a ball slightly to the left and slightly to the right on a hill. We want to find the direction where it goes down the hill and go down to find the very bottom, which is where we could have a very least error. After sliding it to the left and the right, we find the downhill direction, and we leave it there.

Doing that over lots of iterations and steps helps all of the values across every feature. When all the weights settle into a minimum, Back-Propagation ends. At that point, the network is performing as well as it possibly can adjust any of those a little bit its behaviour and error.

2.3 Learn from practically run Convolutional Neural Network

Experience tells us the importance of actual testing in elucidating a BlackBox-like technique. Especially for the users who know only input and output can not understand what happens in between, let alone evaluate the performance of the algorithm. (Refer to the case Deepfake Lab in section 1.3)

Therefore, in the process of testing repeatedly, use the research method of control variables, change the input parameters to observe different output results. Thus seek for how additional input data affects the output results under the same conditions. By visually displaying the input and output data, the audience can understand the role and characteristics of the algorithm clearly without dealing with various terms.

No doubt feeding the same input data to algorithms with different parameters will also produce different outputs. But since both the author and the audience are not Deep Learning professionals, this approach is irrelevant.

2.3.1 Preparation

Practical learning about CNN starts with the knowledge of professional platforms and tools. On platforms such as Medium[20] and Kaggle[21], beginners can find practical tutorials. These tutorials are mainly divided into *Operating environment setup* and *Classify pictures by practically deploying ConvNet*[22].

Operating environment setup

In terms of hardware configuration, Deep Learning requires a powerful GPU(Graphics Processing Unit) to perform complex operations. Usually, neural networks need a large amount of video memory and RAM (Random Access Memory). Therefore, a well-configured server is required to perform complex image recognition tasks. Due to the limitation of hardware conditions, I chose to use the free cloud server, Google Colaboratory (Colab for short), which allows users to run python programs and

[20] Medium is a blog publishing platform founded in August 2012 by Twitter co-founders Evan Williams and Biz Stone. The platform has professional and non-professional contributors, as well as hired editors. It is an example of social news reporting. Among them, towards data science is a column dedicated to information science. The blog content includes data analysis, data visualisation and many Deep Learning projects.

[21] Kaggle is a data modelling and data analysis competition platform. Companies and researchers can publish data on it, and statisticians and data mining experts can compete on it to produce the best model.

[22] Indeed, ConvNet can also be used to identify audio, but this article only discusses the application of ConvNet in image recognition.

synchronise files on Google Drive through code. Colab is beginner-friendly, has the same interface and operating habits as Jupyter Notebook[23], and can record the running results.

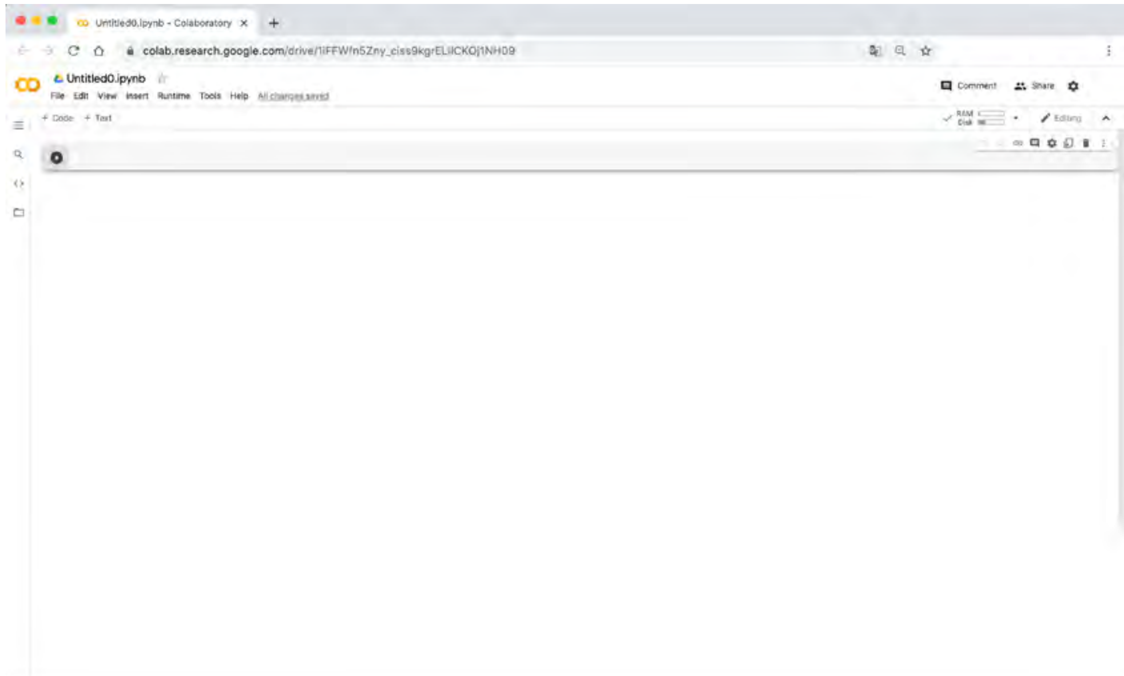


Figure. 34 Snapshot of Google Colab

Classify pictures by practically deploying ConvNet

These tutorials on practical applications often include training pictures and neural network code. Beginners can obtain the code with annotations through the link to the Github[24] repository or individual Jupyter Notebook files(*.ipynb).

[23] The Jupyter Notebook is an open-source web application that allows users to create and share documents that contain live code, equations, visualisations and narrative text.

[24] GitHub is a software source code hosting service platform for version control through Git. Individuals and organisations can establish and store public or private code warehouses. As of 2015, GitHub has more than 28 million registered users and 57 million codebases. It has become the world's largest code storage website and open source community.

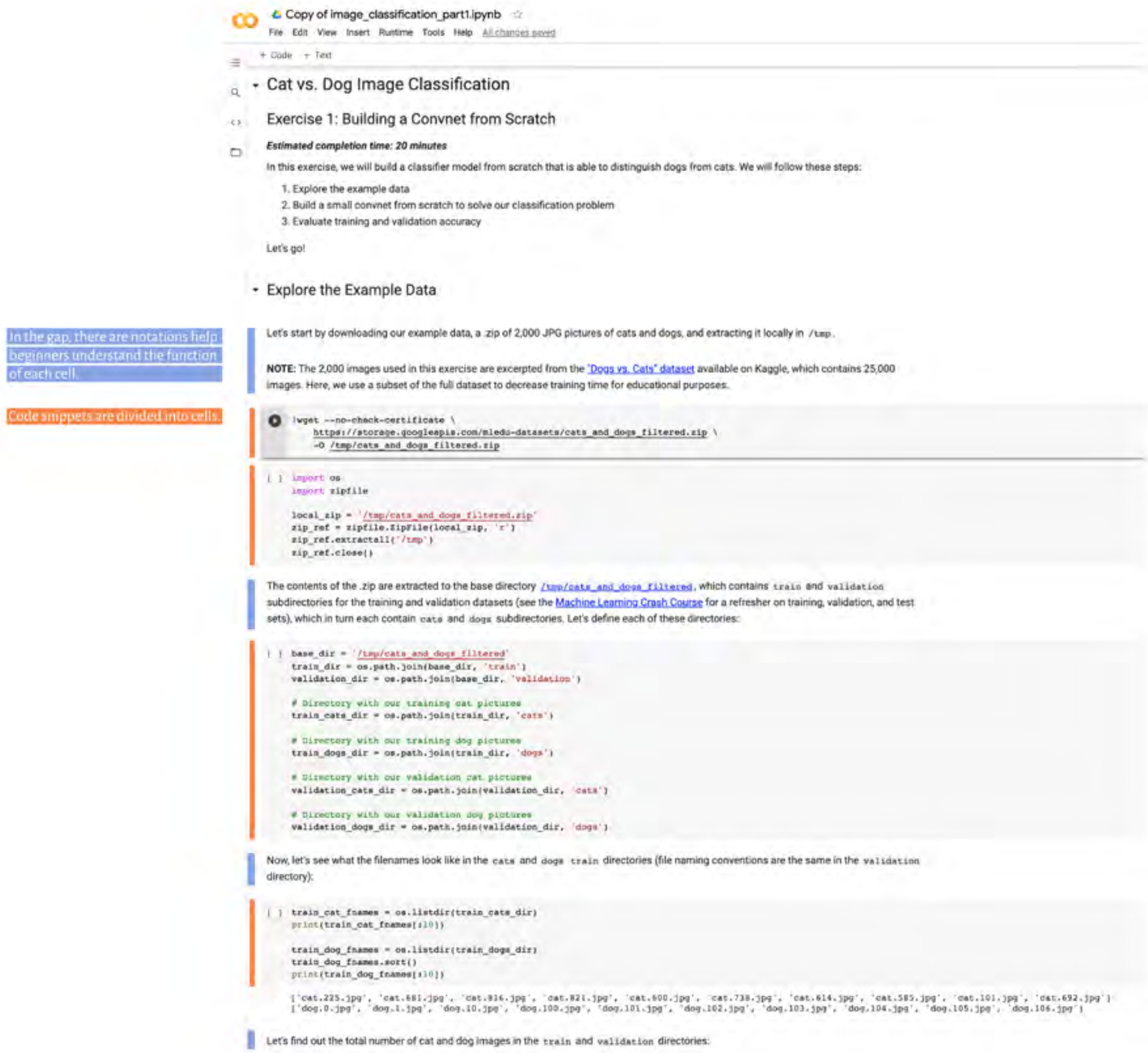


Figure. 35.1 The sample code from the tutorial is shown on Colab

Figure. 35.2 The interaction process on Colab

2.3.2 Image recognition: Cats vs Dogs

The pragmatic programme learning about ConvNets started from the most basic binary classification: image recognition between cats and dogs. The whole process can be divided into three steps: *Preliminary*, *Create and train ConvNet*, and *Validate ConvNet*.

Preliminary

After fixing the computer and having the environment to run the code, we need to prepare datasets of cat and dog images that we can usually find on Kaggle. We need to review these images before using them and improve the quality of the Dataset by eliminating meaningless images (letters, silhouettes, stick figures, etc.) It's as if parents of minors are choosing books for their children with a focus on violence and Gore that might mislead them.

Figure. 36 "Wash" dataset

Create and train ConvNet

Let's code up the architecture. We will stack 3 <convolution + relu + max-pooling> modules.

```
[9]  from tensorflow.keras import layers
     from tensorflow.keras import Model

[10] # Our input feature map is 150x150x3: 150x150 for the image pixels, and 3 for
     # the three color channels: R, G, and B
     img_input = layers.Input(shape=(150, 150, 3))

     # First convolution extracts 16 filters that are 3x3
     # Convolution is followed by max-pooling layer with a 2x2 window
     x = layers.Conv2D(16, 3, activation='relu')(img_input)
     x = layers.MaxPooling2D(2)(x)

     # Second convolution extracts 32 filters that are 3x3
     # Convolution is followed by max-pooling layer with a 2x2 window
     x = layers.Conv2D(32, 3, activation='relu')(x)
     x = layers.MaxPooling2D(2)(x)

     # Third convolution extracts 64 filters that are 3x3
     # Convolution is followed by max-pooling layer with a 2x2 window
     x = layers.Conv2D(64, 3, activation='relu')(x)
     x = layers.MaxPooling2D(2)(x)
```

Figure. 37 frame the network from scratch

On top of it, we stick two fully-connected layers. Because we are facing a two-class classification problem, i.e. a binary classification problem, we will end our network with a sigmoid activation, so that the output of our network will be a single scalar between 0 and 1, encoding the probability that the current image is class 1 (as opposed to class 0).

Let's summarise the model architecture:

```
model.summary()

Model: "functional_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 150, 150, 3)]     0
_________________________________________________________________
conv2d (Conv2D)              (None, 148, 148, 16)      448
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 74, 74, 16)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 72, 72, 32)        4640
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 36, 36, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 34, 34, 64)        18496
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 17, 17, 64)        0
_________________________________________________________________
flatten (Flatten)            (None, 18496)             0
_________________________________________________________________
dense (Dense)                (None, 512)               9470464
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 513
=================================================================
Total params: 9,494,561
Trainable params: 9,494,561
Non-trainable params: 0
_________________________________________________________________
```

Figure. 38 architecture review

Neural networks are trained by presenting them with batches of images, each of them with a label identifying the true nature of the image (either cat or dog in our case). For each image, the prediction of the network is compared with the corresponding label, and the distance between the predictions of the network and the truth is evaluated for the whole batch. Then, after the processing of the batch, the network parameters are changed in a way to minimise this distance, therefore improving the prediction capability of the network.

Next, we'll configure the specifications for model training. We will train our model with the `binary_crossentropy` loss because it's a binary classification problem and our final activation is a sigmoid.

During training, we will monitor classification accuracy.

```
history = model.fit_generator(
      train_generator,
      steps_per_epoch=100,  # 2000 images = batch_size * steps
      epochs=15,
      validation_data=validation_generator,
      validation_steps=50,  # 1000 images = batch_size * steps
      verbose=2)

WARNING:tensorflow:From <ipython-input-15-91d346d1b720>:7: Model.fit_generator (from tensorflow.python.keras.engine.training) is deprecated and will be
Instructions for updating:
Please use Model.fit, which supports generators.
Epoch 1/15
100/100 - 8s - loss: 0.7528 - acc: 0.5760 - val_loss: 0.6353 - val_acc: 0.6800
Epoch 2/15
100/100 - 8s - loss: 0.6176 - acc: 0.6670 - val_loss: 0.6126 - val_acc: 0.6340
Epoch 3/15
100/100 - 8s - loss: 0.5248 - acc: 0.7380 - val_loss: 0.6303 - val_acc: 0.6320
Epoch 4/15
100/100 - 8s - loss: 0.4436 - acc: 0.8030 - val_loss: 0.8882 - val_acc: 0.6010
Epoch 5/15
100/100 - 8s - loss: 0.3666 - acc: 0.8460 - val_loss: 0.5933 - val_acc: 0.7370
Epoch 6/15
100/100 - 8s - loss: 0.2694 - acc: 0.8820 - val_loss: 0.6025 - val_acc: 0.7460
Epoch 7/15
100/100 - 8s - loss: 0.1934 - acc: 0.9235 - val_loss: 0.8794 - val_acc: 0.7020
Epoch 8/15
100/100 - 8s - loss: 0.1567 - acc: 0.9415 - val_loss: 0.9843 - val_acc: 0.7180
Epoch 9/15
100/100 - 8s - loss: 0.0997 - acc: 0.9630 - val_loss: 1.0311 - val_acc: 0.7270
Epoch 10/15
100/100 - 8s - loss: 0.0694 - acc: 0.9805 - val_loss: 1.6975 - val_acc: 0.6740
Epoch 11/15
100/100 - 8s - loss: 0.0779 - acc: 0.9765 - val_loss: 1.1506 - val_acc: 0.7260
Epoch 12/15
100/100 - 8s - loss: 0.0490 - acc: 0.9865 - val_loss: 1.5595 - val_acc: 0.7270
Epoch 13/15
100/100 - 8s - loss: 0.0402 - acc: 0.9905 - val_loss: 1.9012 - val_acc: 0.6960
Epoch 14/15
100/100 - 8s - loss: 0.0484 - acc: 0.9835 - val_loss: 1.8053 - val_acc: 0.7210
Epoch 15/15
100/100 - 8s - loss: 0.0368 - acc: 0.9895 - val_loss: 1.8602 - val_acc: 0.7350
```

Figure. 39 Train algorithm

To get a feel for what kind of features our convnet has learned, one fun thing to do is to visualise how an input gets transformed as it goes through the convnet.

Let's pick a random cat or dog image from the training set, and then generate a figure where each row is the output of a layer, and each image in the row is a specific filter in that output feature map. Rerun this cell to generate intermediate representations for a variety of training images.

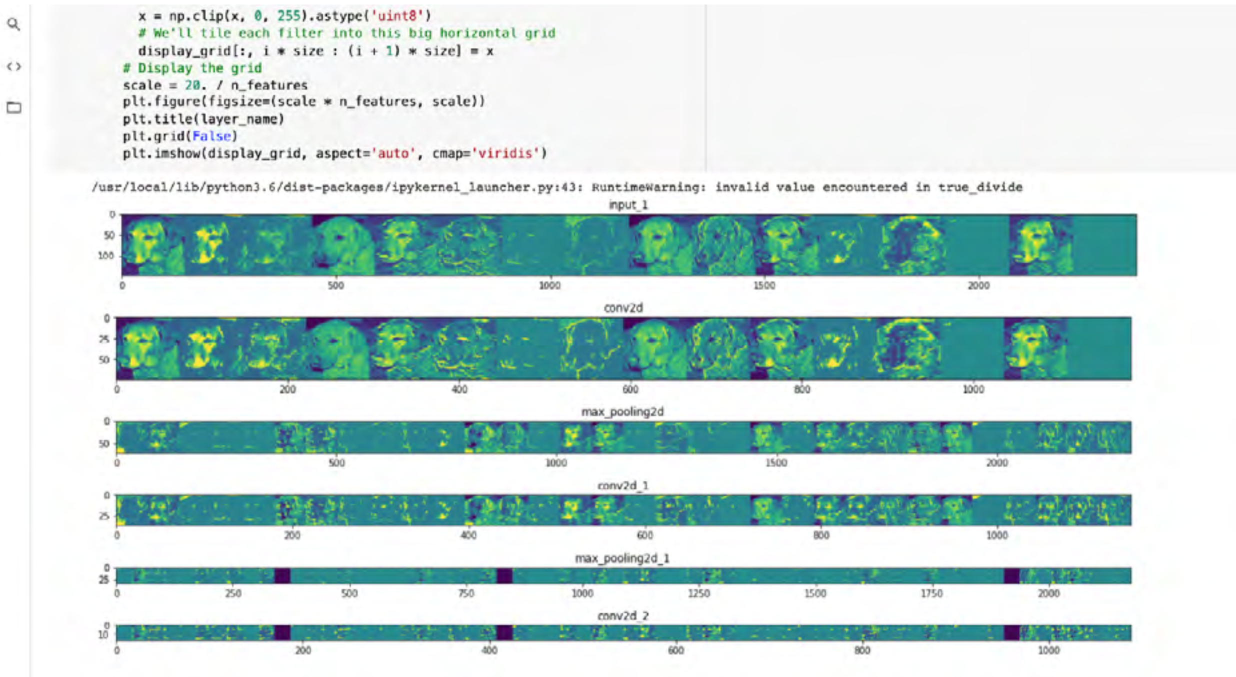


Figure. 40 Intermediate representations of training images

As you can see, we go from the raw pixels of the images to increasingly abstract and compact representations. The representations downstream start highlighting what the network pays attention to, and they show fewer and fewer features being "activated"; most are set to zero. This is called "sparsity[25]". These representations carry less information about the original pixels of the image increasingly, but increasingly refined information about the class of the image. We can think of a convnet (or a deep network in general) as an information distillation pipeline.

Validate ConvNet

In this step, we are testing the accuracy and loss of the trained model. Let's plot the training & validation accuracy and loss as collected during training.

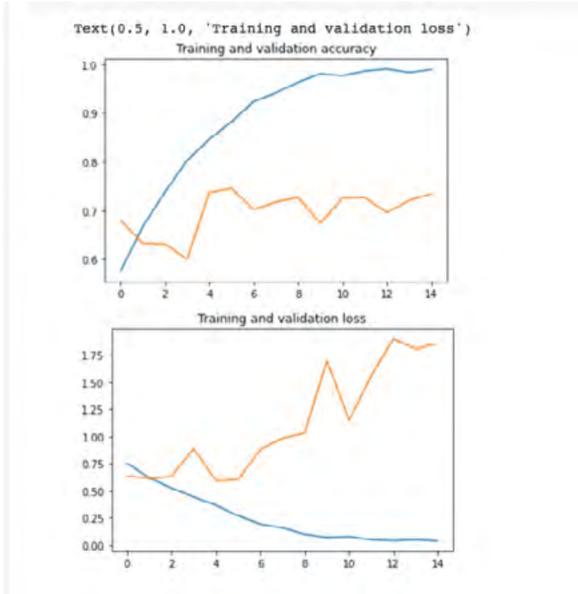


Figure. 41 Accuracy and loss of the trained model

As you can see, we are overfitting[26] like it's getting out of fashion. Our training accuracy (in blue) gets close to 100% (!) while our validation accuracy (in orange) stalls as 70%. Our validation loss reaches its minimum after only five epochs.

Since we have a relatively small number of training examples (2000), overfitting should be our number one

[25] Representation sparsity is a key feature of deep learning.

[26] Overfitting refers to a model that models the training data too well.

concern. Overfitting happens when a model exposed to too few examples learns patterns that do not generalise to new data, i.e. when the model starts using irrelevant features for making predictions.

For instance, if you, as a human, only see three images of people who are lumberjacks and three pictures of people who are sailors, and among them, the only person wearing a cap is a lumberjack, you might start thinking that wearing a cap is a sign of being a lumberjack as opposed to a sailor. You would then make a pretty lousy lumberjack/sailor classifier. Thinking about the “Checkerboard deduction”, we met a similar problem; the “algorithm” is not able to recognise the rotated O because it was not fed enough data in the training step.

Overall

From designing the structure of ConvNet, training it with a suitable dataset, and finally checking the accuracy and loss of recognition, we can find that the workflow of designing a computer program and enriching its ability to recognise objects is not linear. Instead, a loop validates the results of the trained model to determine whether you need to go back to step 1 to reset the parameters or step 2 to train with another Dataset.

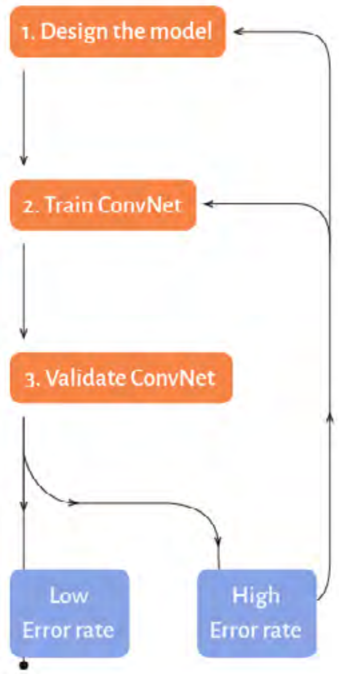


Figure. 42 Workflow of create a effective ConvNet

2.3.3 Insights from practice

For the research on visualising Machine Learning related technologies, the charm of the algorithm is genuinely felt after the program is actually run, rather than limited to the various ConvNet architectures from classic papers. Although my initial attempt was also based on a hands-on online tutorial, and accompanied by a large number of operational bugs, a simple convolutional neural network was stumblingly trained.

In the process of playing with the code, I found that the program is highly parameterised. It understands the content of the picture by creating categories for a series of very subtle features such as colour, edge shape, and material. But the disadvantage is that it does not have a midway correction mechanism. If the model is not adequately designed or there is a problem with the dataset annotation, the person who executes the program needs to be able to find the issue after the training is completed.

From the perspective of a visual designer, I think such an interdisciplinary attempt is worthwhile. The actual running of the program can help the visual work of the storyteller to remove the fog of theoretical knowledge, understanding the principle. To be able to use symbolic and other rhetorical translation of “professional information” to help users through the usual way to understand what’s going on of ConvNet.

INFORMATION

VISUALISATION

3.1 Final user

The media have mentioned concepts such as machine learning and deep learning with a high frequency. To understand these professional terms, it means that people need to spend time reading a lot of articles and understanding mathematical formulas that are complicated. So most people give up at the beginning, but these new technologies are closely related to our technological life.

In the previous chapters (2,2,1), we analysed the audience portraits of ConvNet in online media. In this section, I want to explore further the target audience of this work based on the last three groups of people.

1. Lay user

These are individuals who typically have no prior knowledge about ConvNet, and may not have a technical background. They simply use AI-powered devices and consumer applications.

3.1.1 Why do lay users need to be informed?

Let's be more specific because the general public does not have expertise in computer science or data science; they can't perceive these mature applications of machine learning related technologies in their lives; such as the cooking videos on youtube; after likes, the homepage continuously pushes the cooking content of famous chefs. And while using Spotify for some time, we like the playlists generated by the system more and more. All of these details will all be summarised into, nowadays technology has improved again, and that’s a good guess!

One of the indispensable steps is often overlooked, which is the process of the user inputting information: because of the mere likes or favourite actions, the backstage of the video website can learn about the user’s interest in this type of video; the same is right for music player software, where favourite and skipped tracks are systematically known after a period of use, so following ways are recommended to suit the taste.

In other words, the process of the user inputting information is actually that Internet service providers are collecting our data for the sake of improving the experience. Therefore, ordinary users need to understand the entire flow and processing of information, and ConvNet recognition image is a very typical machine learning related technology. By being informed of the workflow of ConvNet, users can quite understand the basic concept of “input information” -> “processing information” -> “output information” in machine learning related technologies.

The pace of technology development is exponential; many algorithm errors have been looming in recent years. As outsiders who do not understand the decision-making process in the black box, they cannot evaluate and give suggestions for modification; It is a weak recipient under the entire operating system. Thus, understanding such knowledge can arouse their awareness from an individual perspective, and from the overall point of view can give ordinary users the ability to make an opinion.

3.1.2 Carrier selection

Usage

Due to the rapid development of the mobile Internet, people rarely use computers outside of work, and mobile phones have become indispensable in life. As shown in the figure below, we can see that in addition to the sluggish number of people using tablets to access the Internet, the number of mobile phone users and desktop computer users has always been huge. Since the first quarter in 2019, the number of mobile phone users has begun to exceed the number of desktop computer users. Therefore, compared to the traditional computer-side webpage, the mobile terminal as the carrier of the website is closer to the user's habits of using the Internet.

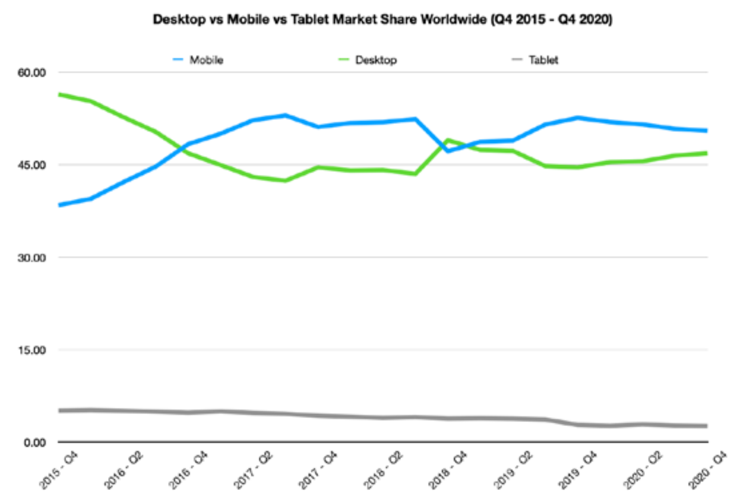


Figure. 43 Desktop vs Mobile vs Tablet Share Worldwide (2015-2020)
Data Source

Volume of information

The big screen can indeed carry more information and details, but the project is positioned as a relatively bright science-oriented visual visualisation website. The limited space of each screen can restrict designers to translate professional information more thoroughly and brightly.

3.1.3 User-oriented content filtering

In the early stages of learning research, the author has gained a lot of professional knowledge about machine learning, whether it is academic or practical. The important thing is that from the beginners at the beginning of the research to the introductory students who have been able to train neural networks under the guidance of computer science experts, the knowledge of ConvNet and related technologies is also richer. Therefore, I can no longer think of myself as an ordinary user with no foundation and ask myself "Do I already understand?" to measure whether my target users can understand this information.

But when introducing ConvNet related technologies (deep learning and machine learning, etc.), in addition to the main introduction of ConvNet image recognition workflow and algorithm errors caused by uneven data sets, it also sets Another layer of information structure, including Glossary, Supplement and About. As shown in Figure n, if a professional term is mentioned in the light-coloured page, the terminologies are underlined to indicate that they are clickable buttons, and the user enters the glossary interface after taping it.

After browsing, click the cross icon in the upper left corner to restore the original light-coloured page. Some nouns involving complicated visualised interpretation are the same.

If the user does not ask for a thorough understanding while browsing, it will not seriously affect the knowledge if the user is inquisitive about the details. This set of information structure effectively filters relatively complex and professional content, which may make some users feel bored and stop browsing the web.

In addition to the explanation of individual underlined words, users can also enter the main menu by tapping the menu bar in the upper right corner, and then click the corresponding icon to enter different additional modules for retrieving.

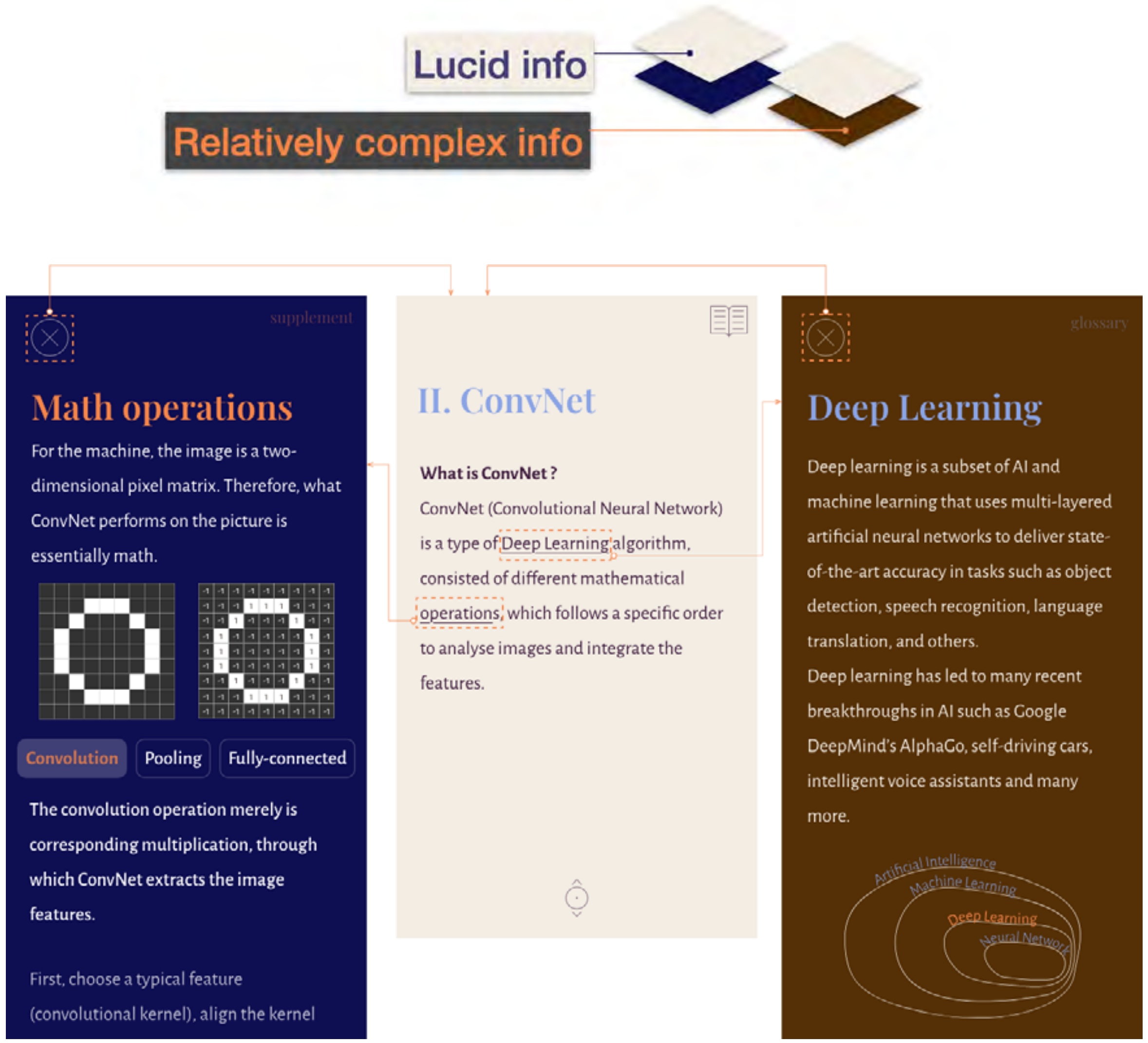


Figure. 44.a. b. Layered information for filtering different users

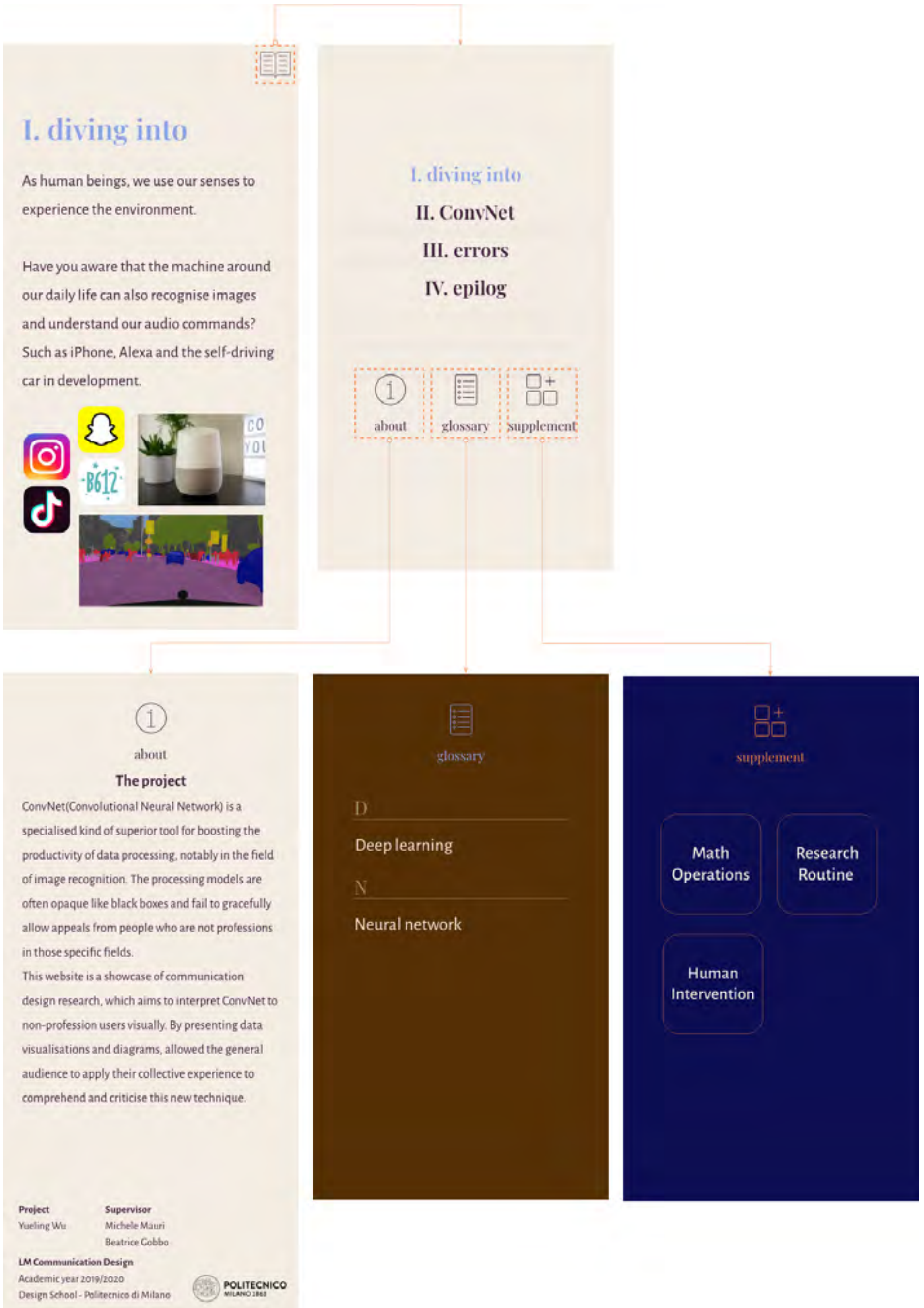
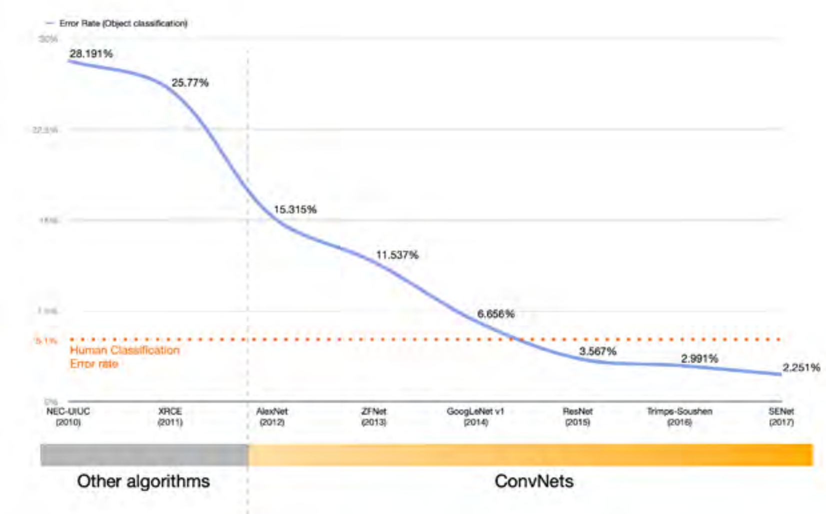


Figure. 45 Path for retrieving the additional modules

3.2 Visualisation

3.2.1 General visualisation

In the process of web design, in addition to using text and small icons to explain this new technology, many simple visualisations were also deployed—a simple histogram, as shown in the figure below. Under normal circumstances, data such as numerical probability should be presented by a line chart, but confined to the long and narrow screen of the mobile phone; the author needs to deform the line chart on the left and fit it into space on the right. Therefore, a horizontal histogram is used to present the ConvNet recognition error rate that is rapidly decreasing year by year.



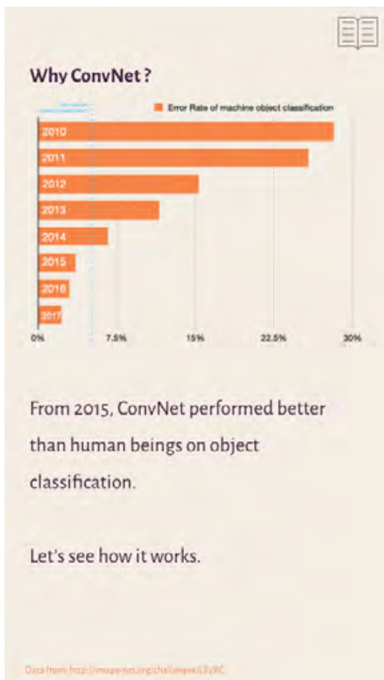


Figure. 46 a.b. Transformation of same visualisation

A similar example is shown in the figure below. In general, the flow chart is stretched horizontally. Still, because of the interactive way of scrolling on the mobile phone screen, the extended vertical process analysis diagram can also effectively display information.

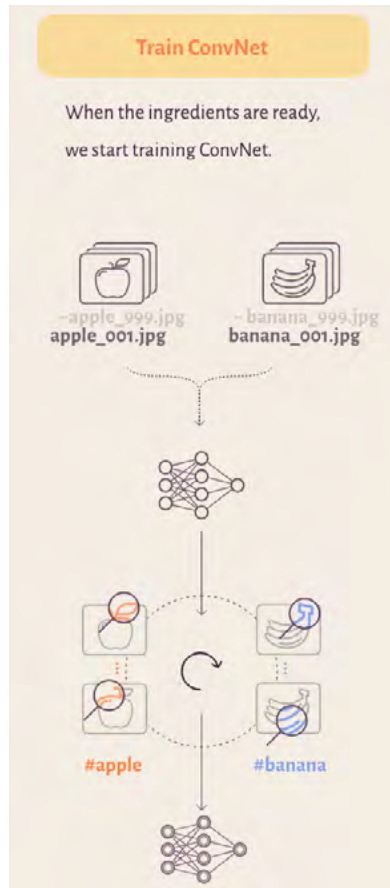


Figure. 47 Vertical workflow diagram

3.2.2 Experiments and visualisations

The two main visualisations revolve around the core problem of the real project: *How does the level of training affect the recognition accuracy of ConvNet?* And *How does the proper sampled training data set affect the recognition accuracy of ConvNet?*
The workflow for making these two visualisations is roughly the same:

1. Prepare datasets (i.e. pictures) according to the experimental goals, and divide them into training data sets and validation data sets.

2. Design ConvNet according to the identification content of the experiment

3. Use different training data sets to train ConvNet, and save the model and number

4. Use each trained model in the previous step to identify the same verification data set and record the experiment results

5. Design visualisation by observing experimental data

How does the level of training affect the recognition accuracy of ConvNet?
Since this is the first experiment when the user is browsing the web, the content of the experiment is determined as follows: Train ConvNet to recognise five common fruits (green apple, banana, honeydew, red apple and star fruit). The variables of the experiment are data sets of different sizes, namely.

- Group_5%: 50 pieces of each fruit
(a total of 250 images in the training data set);

- Group_10%: 100 pieces of each fruit
(a total of 500 images in the training data set);

- Group_25%: 250 pieces of each fruit
(a total of 1250 images in the training data set);

- Group_50%: 500 pieces of each fruit
(a total of 2500 images in the training data set);

- Group_100%: 1000 pieces of each fruit
(a total of 5000 images in the training data set)

After testing ConvNet with the validation data set and recording the accuracy, we randomly selected a group of fruits for five validating images and plotted several histograms showing the recognition accuracy, as shown in Figure 48 on the right.

An interesting finding is obtained: when the training data set contains a small amount of data (5%, 10%), the trained ConvNet often gives multiple recognition results based on an object; on the contrary, when the training data set contains a large amount of data When (100%), trained ConvNet will output fewer recognition results and mostly accurate. Therefore, we recorded both the value validated as the correct fruit and the deal validated as the wrong fruit and used the former to draw a line graph(As Figure 49. a.b.c).

Truth	(50 each group) 5%	(100 each group) 10%	(250 each group) 25%	(500 each group) 50%	(1000 each group) 100%
Apple_green	15	0	70	100	100
Banana	74	96	100	100	100
Melon	24	27.5	5	90	100
Apple_red	94	99	98	100	100
Starfruit	26	91	99	100	100

Figure. 49 a.

Figure. 48 Bar charts showing recognition possibility
(From top to bottom: 5%, 10%, 25%, 50%,100%)

Truth	recognition result	(50 each group) 5%	(100 each group) 10%	(250 each group) 25%	(500 each group) 50%	(1000 each group) 100%
Apple_green	Melon	20	3	2	0	0
	Banana	30	7	3	0	0
	Apple_green	15	0	70	100	100
	Starfruit	35	90	25	0	0
Banana	Banana	74	96	100	100	100
	Melon	11	2	0	0	0
	Starfruit	15	2	0	0	0
Melon	Starfruit	46	34	20	10	0
	Apple_green	3	0	0	0	0
	Melon	24	27.5	5	90	100
	Banana	27	38.5	75	0	0
Apple_red	Banana	1	0	1	0	0
	Apple_red	94	99	98	100	100
	Starfruit	5	1	1	0	0
Starfruit	Starfruit	26	91	99	100	100
	Apple_green	5	0	0	0	0
	Melon	64	4	1	0	0
	Banana	5	5	0	0	0

Figure. 49 b.

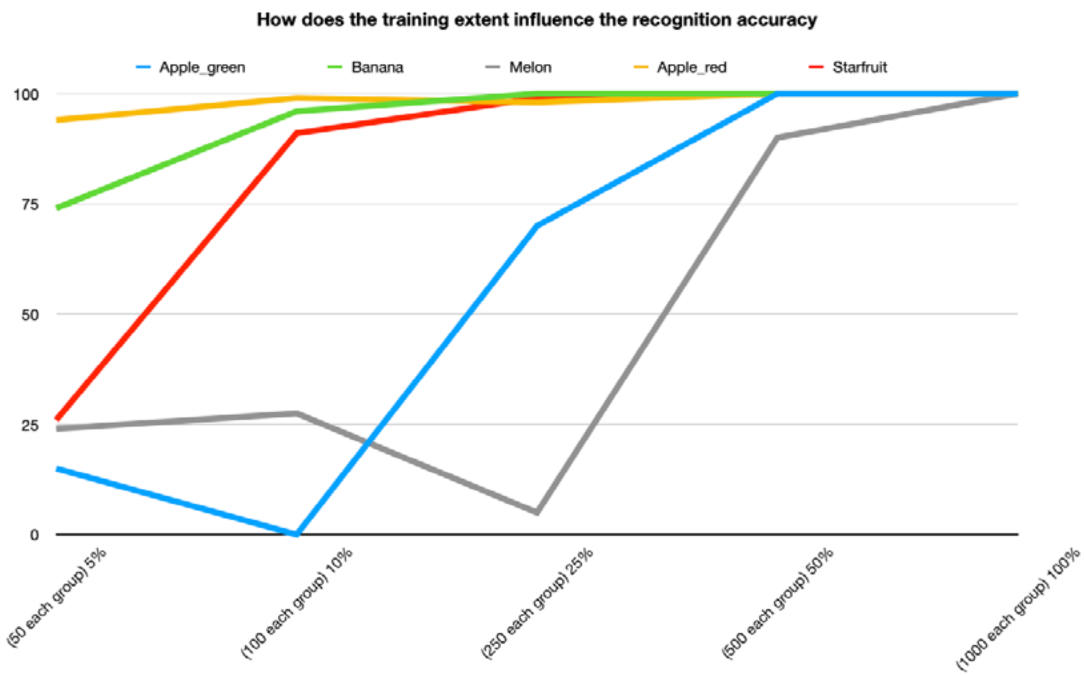


Figure. 49 c.

The line graph shows that from a general trend, the larger the training image data set of ConvNet, the higher the recognition accuracy. There are also instabilities. For example, the green apples and melons represented by the blue lines and grey lines declined after an inevitable rise. This shows that there are also accidents in the verification pictures.

Inspired by the low training level ConvNet, the author presents the stale recognition results by superimposing fruit pictures with different degrees of transparency.

Figure. 50 interface of first visualisation

However, due to the complexity of the visual model, the author decided to put it on the second layer of the page, that is, users need to touch the screen to browse to it to avoid confusion. Therefore, the primary visual interface is relatively straightforward at a glance. As shown below

Figure. 51.a interfaces of the first visualisation

The page is mainly composed of four parts: the progress bar at the top shows the training level of the ConvNet corresponding to the visualisation on this page; the second part is to stack training images to make users have an intuitive experience of the number of training images; the third part is to hide the visualisation After the user taps on the entrance, a complex model will be presented; the fourth part is a simple histogram, showing the value of ConvNet on this page that correctly recognises the fruit.

Jump out of the interface, and the user can quickly swipe the screen horizontally to switch the training level to get a dynamic experience of the bottom bar graph.

Figure. 51.b interfaces of the first visualisation: the hidden space (10%, 20%, 50%))

How does the proper sampled training data set affect the recognition accuracy of ConvNet?

We are assuming that after browsing the first simple fruit recognition experiment, the user has some basic knowledge about ConvNet to recognise pictures. Therefore, the second experiment method will be more complicated, and the content of recognition will be more sensitive: skin colour and gender.

As a beginner, the data set is the only experimental variable that we can precisely control. Like the first experiment, we need to prepare the training data set before training ConvNet carefully. As shown in the figure, there are five groups as follows:

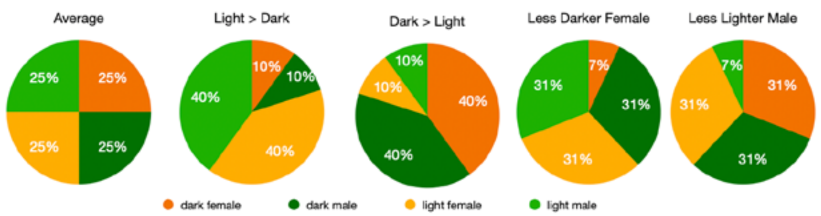


Figure. 52 Bar chart showing the composition of five training datasets

	dark female	dark male	light female	light male
Average	200	200	200	200
① Lighter > Darker	50	50	200	200
② Darker > Lighter	200	200	50	50
③ Less Darker Female	35	155	155	155
④ Less Lighter Male	155	155	155	35

Table. 6 Amount of images of each dataet under different catagory

The first evenly distributed data set is used as a control group to confirm that the ConvNet model itself has no problems. The following four data sets test the effects of the overall skin colour and sample unevenness of a single group on the recognition results.

The model of this experiment is VGG16 pre-trained by ImageNet. Although the lighter male cannot be marked as daily objects, the pre-trained model can extract the features from the data more quickly.

Similar to the first experiment, after training the same ConvNet model with five sets of training data sets, number and save the model. The next step is to test the performance of ConvNet with the validation data set. The results are shown in the following table:

recall TP/(TP+FP)	Average: df=dm=lf=lm=200img	① Light > Dark: 500total 10%+10% df=dm=50img 40%+40% lf=lm=200img	② Dark > Light: 500total 10%+10% lf=lm=50img 40%+40% df=dm=200img	③ Poor df: 500 img total 7% df=35 31%=dm=lf=lm=155img	④ Poor lm: 500 img total 7% lm=35 31%=lf=dm=df=155img
dark female	86%	78%	87%	59%	88%
dark male	89%	56%	76%	75%	76%
light female	87%	87%	74%	89%	89%
light male	84%	82%	66%	68%	57%

Table. 7 Validation results

The percentage in the table is the correct recognition rate (also known as the recall value), such as 78% from group_01 Light> Dark. In the validation process, we use 160 pictures (640 in total) under each category for validation test. Of the 160 images classified by Darker Female, 78% (i.e. 124 images) Darker Female was correctly identified.

By observing the table data, we can find that the correct recognition rate of the four categories in the average sampled control group is equivalent; the correct recognition rate of the two Darker categories in group_01 Light> Dark is significantly lower than the recognition rate of the two Lighter categories; group_02 Dark> Light The situation is the opposite; in the last two groups, Darker Female and Lighter Male, which account for only 7% of their respective training data sets, are the lowest. It is confirmed that the experimental results are consistent with the hypothesis: ConvNet trained with unevenly sampled data sets has a low correct recognition rate for the lesser data types in the data set. However, it has not been proved that there is a linear

relationship between the proportion of each category in the training data set and the correct recognition rate of the category after verification.

Similar to the previous experiment, in the process of validating ConvNet, we also randomly selected a set of pictures for 5-word verification and plotted the recognition probability histogram, as shown below. Multiple recognition results were found in both the

Figure. 53 a. Result of average group

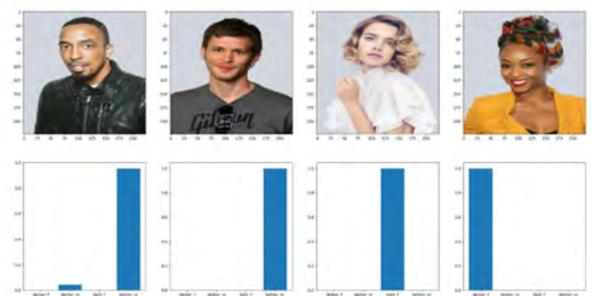


Figure. 53 b. Result of Lighter > Darker group

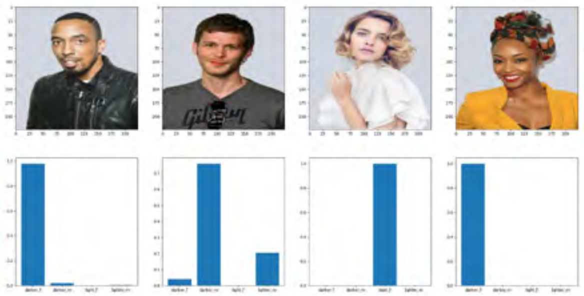


Figure. 53 c. Result of Darker > Lighter group

Figure. 53 d. Result of Less Darker Female group

Figure. 53 e. Result of Less Lighter Male group

02 and 03 groups. So I output all the recognition probabilities of ConvNet for the four categories in batches, and made them in the form of a table, as shown below.

By observing the table, we can find that there are some

01.txt

```
img:df_01.png, darker_f:0.00000, darker_m:0.00000, light_f:0.99999, lighter_m:0.00001
img:df_010.png, darker_f:0.99999, darker_m:0.00000, light_f:0.00000, lighter_m:0.00000
img:df_011.png, darker_f:1.00000, darker_m:0.00000, light_f:0.00000, lighter_m:0.00000
img:df_012.png, darker_f:0.99999, darker_m:0.00001, light_f:0.00000, lighter_m:0.00000
img:df_013.png, darker_f:0.98466, darker_m:0.01396, light_f:0.00138, lighter_m:0.00000
img:df_014.png, darker_f:0.00000, darker_m:0.99986, light_f:0.00000, lighter_m:0.00014
img:df_015.png, darker_f:0.99966, darker_m:0.00031, light_f:0.00003, lighter_m:0.00000
img:df_02.png, darker_f:1.00000, darker_m:0.00000, light_f:0.00000, lighter_m:0.00000
img:df_03.png, darker_f:1.00000, darker_m:0.00000, light_f:0.00000, lighter_m:0.00000
img:df_04.png, darker_f:0.91040, darker_m:0.08957, light_f:0.00002, lighter_m:0.00000
img:df_05.png, darker_f:0.99848, darker_m:0.00000, light_f:0.00152, lighter_m:0.00000
img:df_06.png, darker_f:0.99999, darker_m:0.00000, light_f:0.00001, lighter_m:0.00000
img:df_07.png, darker_f:0.99995, darker_m:0.00001, light_f:0.00003, lighter_m:0.00000
img:df_08.png, darker_f:1.00000, darker_m:0.00000, light_f:0.00000, lighter_m:0.00000
img:df_09.png, darker_f:0.99988, darker_m:0.00012, light_f:0.00000, lighter_m:0.00000
img:dm_01.png, darker_f:0.95111, darker_m:0.04586, light_f:0.00268, lighter_m:0.00034
img:dm_010.png, darker_f:0.96000, darker_m:0.03893, light_f:0.00105, lighter_m:0.00001
img:dm_011.png, darker_f:0.00004, darker_m:0.99219, light_f:0.00002, lighter_m:0.00775
img:dm_012.png, darker_f:0.00064, darker_m:0.98528, light_f:0.00002, lighter_m:0.01406
img:dm_013.png, darker_f:0.00000, darker_m:0.99615, light_f:0.00000, lighter_m:0.00385
img:dm_014.png, darker_f:0.00000, darker_m:0.00141, light_f:0.00000, lighter_m:0.99859
img:dm_015.png, darker_f:0.00001, darker_m:0.95320, light_f:0.00001, lighter_m:0.04678
img:dm_02.png, darker_f:0.97015, darker_m:0.02679, light_f:0.00302, lighter_m:0.00003
img:dm_03.png, darker_f:0.94266, darker_m:0.03358, light_f:0.02348, lighter_m:0.00028
img:dm_04.png, darker_f:0.03696, darker_m:0.03546, light_f:0.73687, lighter_m:0.19070
img:dm_05.png, darker_f:0.94623, darker_m:0.04402, light_f:0.00896, lighter_m:0.00080
img:dm_06.png, darker_f:0.00001, darker_m:0.18629, light_f:0.00006, lighter_m:0.81364
img:dm_07.png, darker_f:0.01498, darker_m:0.72677, light_f:0.03466, lighter_m:0.22359
img:dm_08.png, darker_f:0.99990, darker_m:0.00008, light_f:0.00003, lighter_m:0.00000
img:dm_09.png, darker_f:0.00576, darker_m:0.99014, light_f:0.00001, lighter_m:0.00409
img:lf_01.png, darker_f:0.01356, darker_m:0.00000, light_f:0.98596, lighter_m:0.00048
img:lf_010.png, darker_f:0.00000, darker_m:0.00000, light_f:1.00000, lighter_m:0.00000
img:lf_011.png, darker_f:0.00000, darker_m:0.00000, light_f:0.37286, lighter_m:0.62714
img:lf_012.png, darker_f:0.00000, darker_m:0.00000, light_f:1.00000, lighter_m:0.00000
img:lf_013.png, darker_f:0.00000, darker_m:0.00000, light_f:1.00000, lighter_m:0.00000
img:lf_014.png, darker_f:0.00004, darker_m:0.00000, light_f:0.99996, lighter_m:0.00000
img:lf_015.png, darker_f:0.00000, darker_m:0.00000, light_f:0.03256, lighter_m:0.96744
img:lf_02.png, darker_f:0.00000, darker_m:0.00000, light_f:1.00000, lighter_m:0.00000
img:lf_03.png, darker_f:0.00000, darker_m:0.00000, light_f:0.99986, lighter_m:0.00014
img:lf_04.png, darker_f:0.00000, darker_m:0.00000, light_f:1.00000, lighter_m:0.00000
img:lf_05.png, darker_f:0.00000, darker_m:0.00000, light_f:1.00000, lighter_m:0.00000
img:lf_06.png, darker_f:0.00000, darker_m:0.00000, light_f:0.85075, lighter_m:0.14925
img:lf_07.png, darker_f:0.00000, darker_m:0.00000, light_f:0.99973, lighter_m:0.00027
img:lf_08.png, darker_f:0.00080, darker_m:0.00000, light_f:0.99918, lighter_m:0.00001
img:lf_09.png, darker_f:0.00000, darker_m:0.00000, light_f:1.00000, lighter_m:0.00000
img:lm_01.png, darker_f:0.00000, darker_m:0.03278, light_f:0.00000, lighter_m:0.96722
img:lm_010.png, darker_f:0.00000, darker_m:0.00781, light_f:0.00072, lighter_m:0.99147
img:lm_011.png, darker_f:0.00000, darker_m:0.00000, light_f:0.00000, lighter_m:1.00000
img:lm_012.png, darker_f:0.00003, darker_m:0.00026, light_f:0.99114, lighter_m:0.00857
img:lm_013.png, darker_f:0.00000, darker_m:0.00000, light_f:0.00000, lighter_m:1.00000
img:lm_014.png, darker_f:0.02181, darker_m:0.10042, light_f:0.13463, lighter_m:0.74315
img:lm_015.png, darker_f:0.00000, darker_m:0.00000, light_f:0.00000, lighter_m:1.00000
img:lm_02.png, darker_f:0.00000, darker_m:0.00230, light_f:0.00000, lighter_m:0.99770
img:lm_03.png, darker_f:0.00000, darker_m:0.00000, light_f:0.00000, lighter_m:1.00000
img:lm_04.png, darker_f:0.00000, darker_m:0.75405, light_f:0.00000, lighter_m:0.24595
img:lm_05.png, darker_f:0.00000, darker_m:0.00052, light_f:0.00025, lighter_m:0.99922
img:lm_06.png, darker_f:0.00000, darker_m:0.00000, light_f:0.00000, lighter_m:1.00000
img:lm_07.png, darker_f:0.00000, darker_m:0.32628, light_f:0.00000, lighter_m:0.67372
img:lm_08.png, darker_f:0.00000, darker_m:0.00422, light_f:0.00008, lighter_m:0.99570
img:lm_09.png, darker_f:0.00007, darker_m:0.00025, light_f:0.99895, lighter_m:0.00073
```

validate_img	dark_f	dark_m	light_f	light_m
df_01.png	0.00	0.00	1.00	0.00
df_010.png	1.00	0.00	0.00	0.00
df_011.png	1.00	0.00	0.00	0.00
df_012.png	1.00	0.00	0.00	0.00
df_013.png	0.98	0.01	0.00	0.00
df_014.png	0.00	1.00	0.00	0.00
df_015.png	1.00	0.00	0.00	0.00
df_02.png	1.00	0.00	0.00	0.00
df_03.png	1.00	0.00	0.00	0.00
df_04.png	0.91	0.09	0.00	0.00
df_05.png	1.00	0.00	0.00	0.00
df_06.png	1.00	0.00	0.00	0.00
df_07.png	1.00	0.00	0.00	0.00
df_08.png	1.00	0.00	0.00	0.00
df_09.png	1.00	0.00	0.00	0.00
dm_01.png	0.95	0.05	0.00	0.00
dm_010.png	0.96	0.04	0.00	0.00
dm_011.png	0.00	0.99	0.00	0.01
dm_012.png	0.00	0.99	0.00	0.01
dm_013.png	0.00	1.00	0.00	0.00
dm_014.png	0.00	0.00	0.00	1.00
dm_015.png	0.00	0.95	0.00	0.05
dm_02.png	0.97	0.03	0.00	0.00
dm_03.png	0.94	0.03	0.02	0.00
dm_04.png	0.04	0.04	0.74	0.19
dm_05.png	0.95	0.04	0.01	0.00
dm_06.png	0.00	0.19	0.00	0.81
dm_07.png	0.01	0.73	0.03	0.22
dm_08.png	1.00	0.00	0.00	0.00
dm_09.png	0.01	0.99	0.00	0.00
lf_01.png	0.01	0.00	0.99	0.00
lf_010.png	0.00	0.00	1.00	0.00
lf_011.png	0.00	0.00	0.37	0.63
lf_012.png	0.00	0.00	1.00	0.00
lf_013.png	0.00	0.00	1.00	0.00
lf_014.png	0.00	0.00	1.00	0.00
lf_015.png	0.00	0.00	0.03	0.97
lf_02.png	0.00	0.00	1.00	0.00
lf_03.png	0.00	0.00	1.00	0.00
lf_04.png	0.00	0.00	1.00	0.00
lf_05.png	0.00	0.00	1.00	0.00
lf_06.png	0.00	0.00	0.85	0.15
lf_07.png	0.00	0.00	1.00	0.00
lf_08.png	0.00	0.00	1.00	0.00
lf_09.png	0.00	0.00	1.00	0.00
lm_01.png	0.00	0.03	0.00	0.97
lm_010.png	0.00	0.01	0.00	0.99
lm_011.png	0.00	0.00	0.00	1.00
lm_012.png	0.00	0.00	0.99	0.01
lm_013.png	0.00	0.00	0.00	1.00
lm_014.png	0.02	0.10	0.13	0.74
lm_015.png	0.00	0.00	0.00	1.00
lm_02.png	0.00	0.00	0.00	1.00
lm_03.png	0.00	0.00	0.00	1.00
lm_04.png	0.00	0.75	0.00	0.25
lm_05.png	0.00	0.00	0.00	1.00
lm_06.png	0.00	0.00	0.00	1.00
lm_07.png	0.00	0.33	0.00	0.67
lm_08.png	0.00	0.00	0.00	1.00
lm_09.png	0.00	0.00	1.00	0.00

Figure. 54 Row data of multiple recognition results

recognition images from different categories but sharing the same identification tag. Such a situation can be visualised in the form of a graph network, and the thickness of the edge connecting the nodes indicates the probability value of the picture being classified into this category. The following figure is based on the experimental results of group_Light > Dark.
Although there are very few Darker Males mixed in the

Figure. 55 Networking style visualisation

Lighter classification, it is evident that the classification results of the two photo clusters related to Darker are more confusing. There is nearly half the number of Darker Males in the Darker Female cluster.

Considering the visual unity, the interface design of the second experiment continues the top and bottom structure of the first experiment: at the top are variable options. By tapping different buttons, users can switch training sets with varying ratios of sampling; in the middle is through regular All training images are displayed by stacking images; the third part is the main image network visualisation.
Similar to the first experiment, by switching training

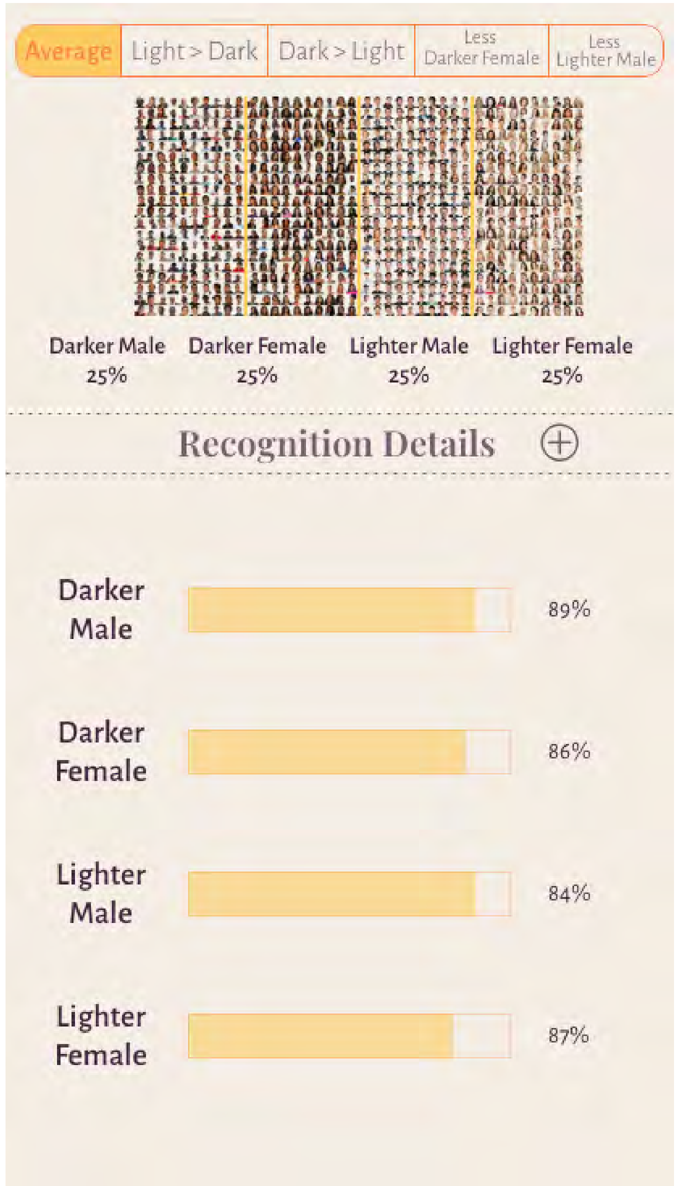


Figure. 56 Recognition results of Group_average

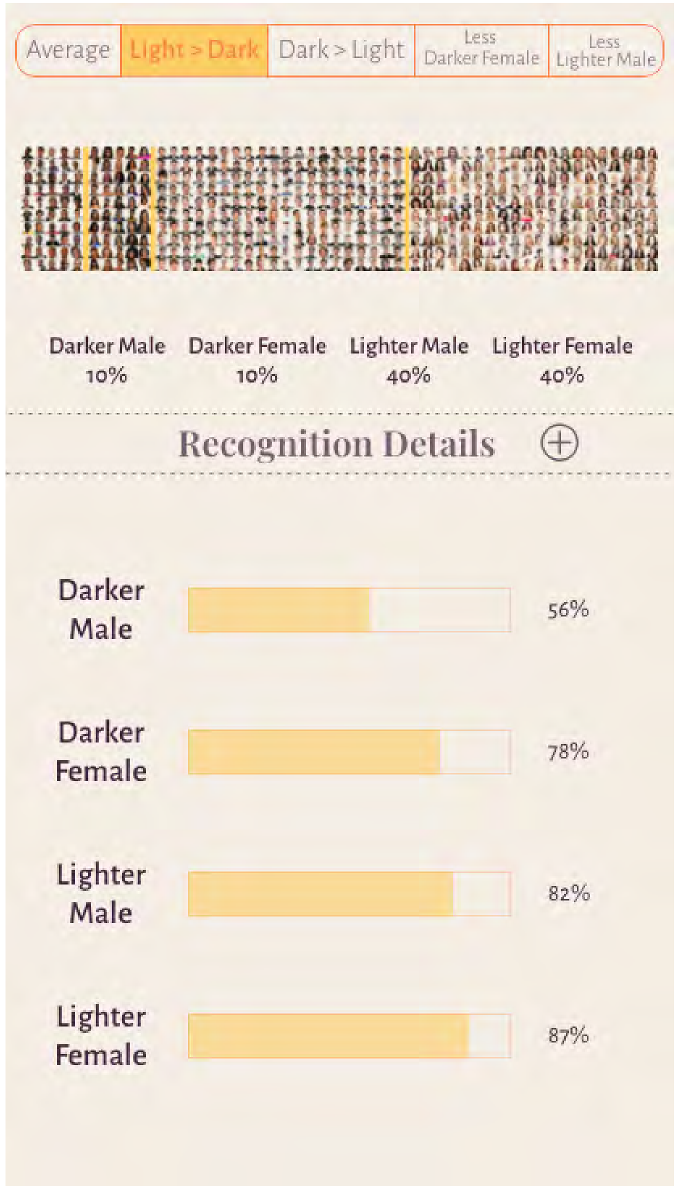


Figure. 57 Recognition results of Group_Light>Dark

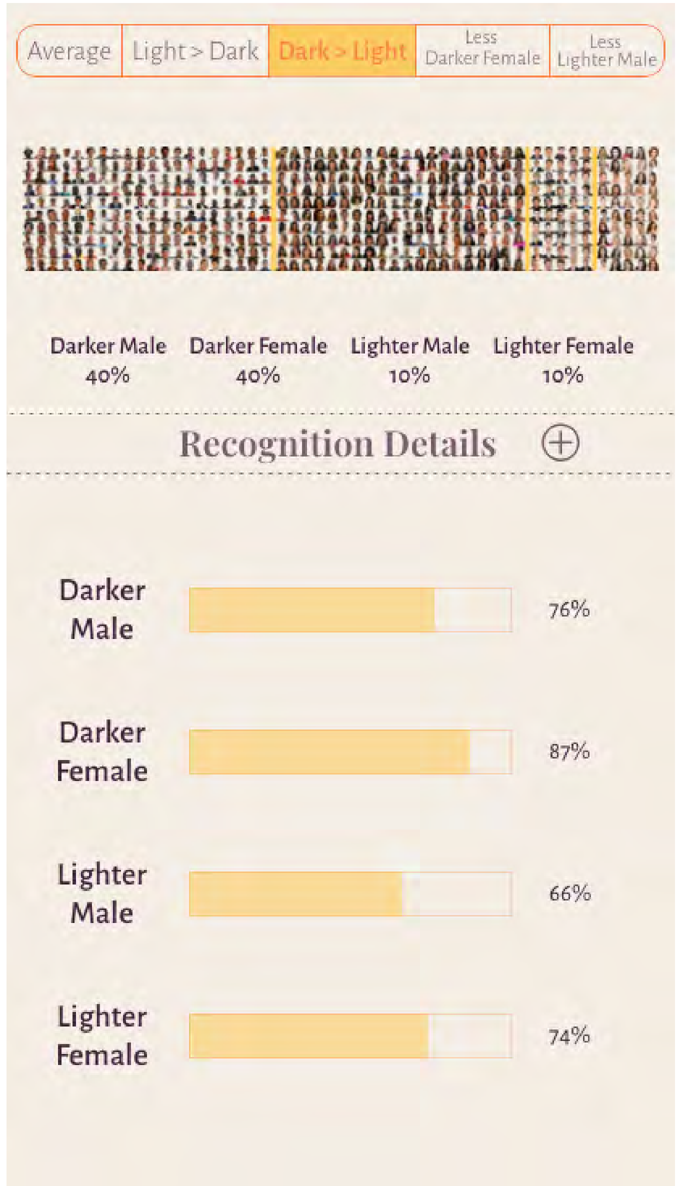


Figure. 58 Recognition results of Group_Dark>Light

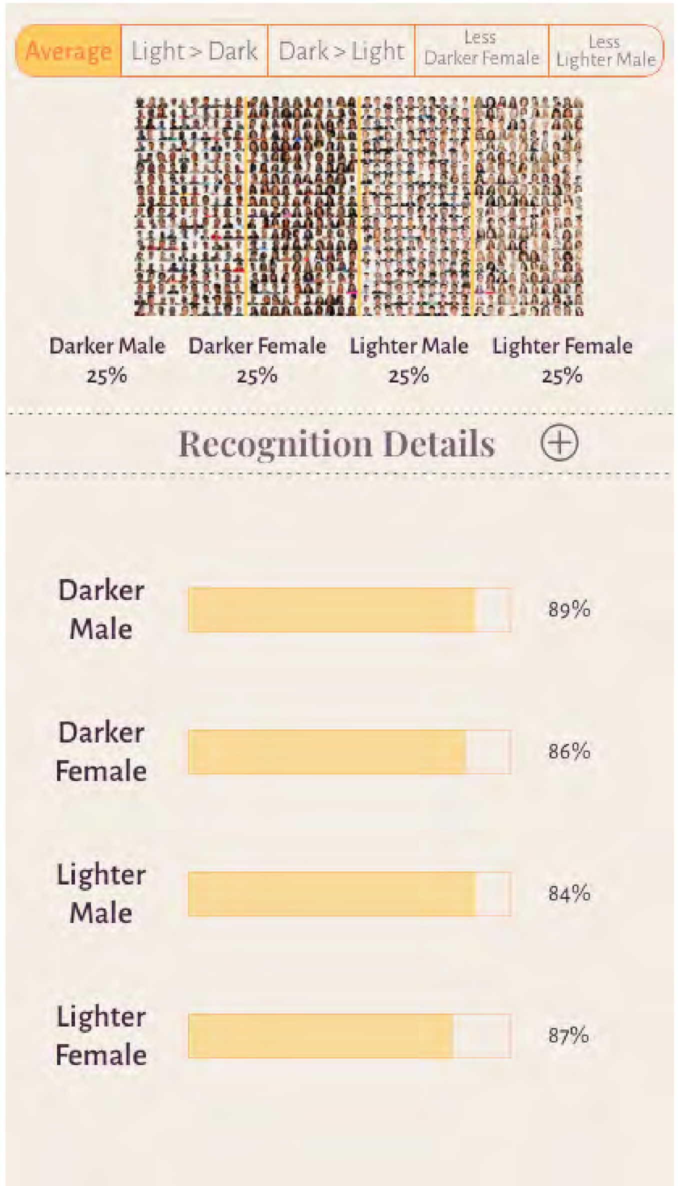


Figure. 56 Recognition results of Group_average

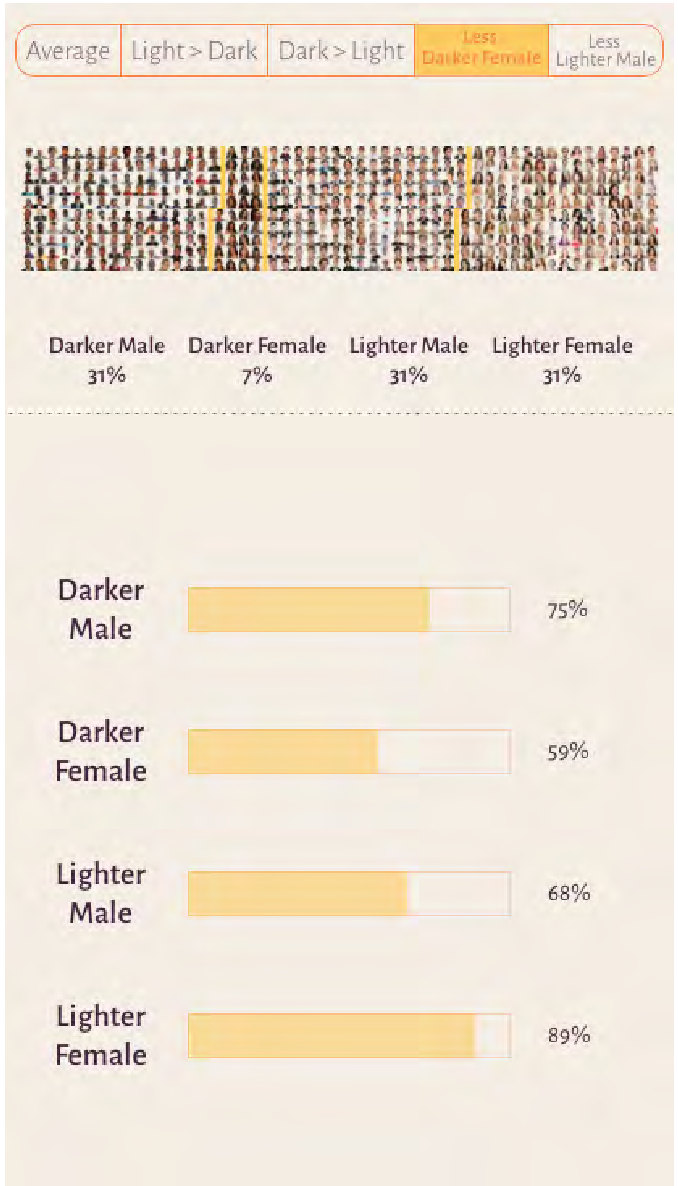


Figure. 59 Recognition results of Group_Less Darker Female

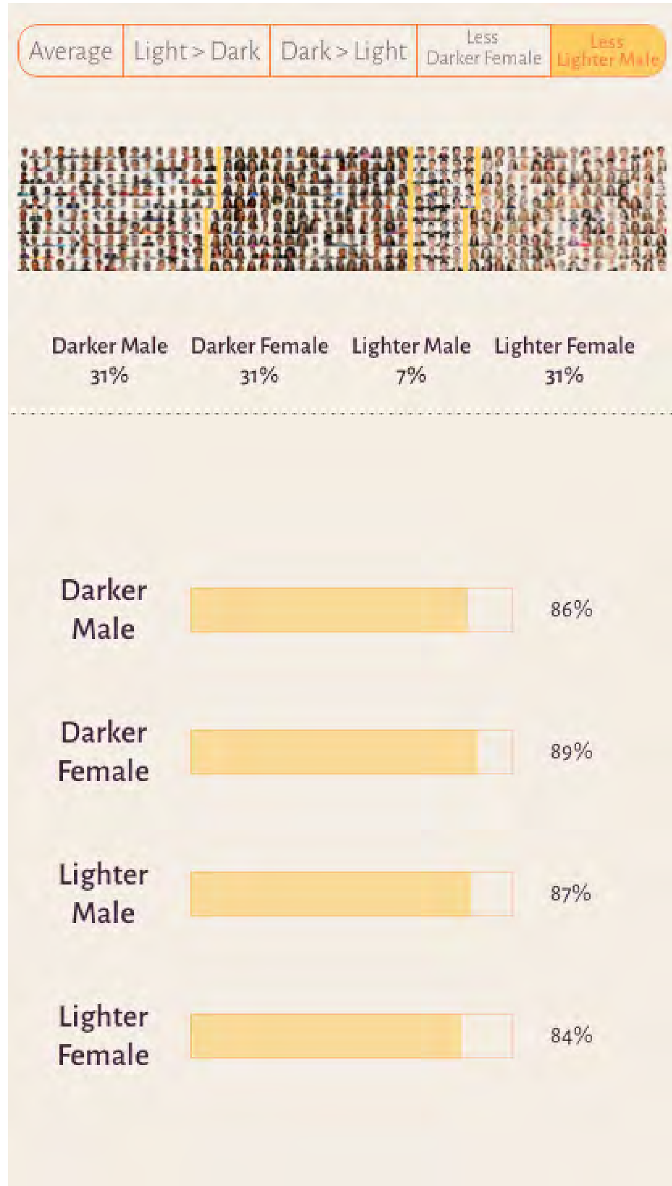


Figure. 60 Recognition results of Group_Less Lighter Male

sets with different sampling ratios, by tracking the relative positions of the same picture and the four classification clusters, users can have a general understanding of the rules.

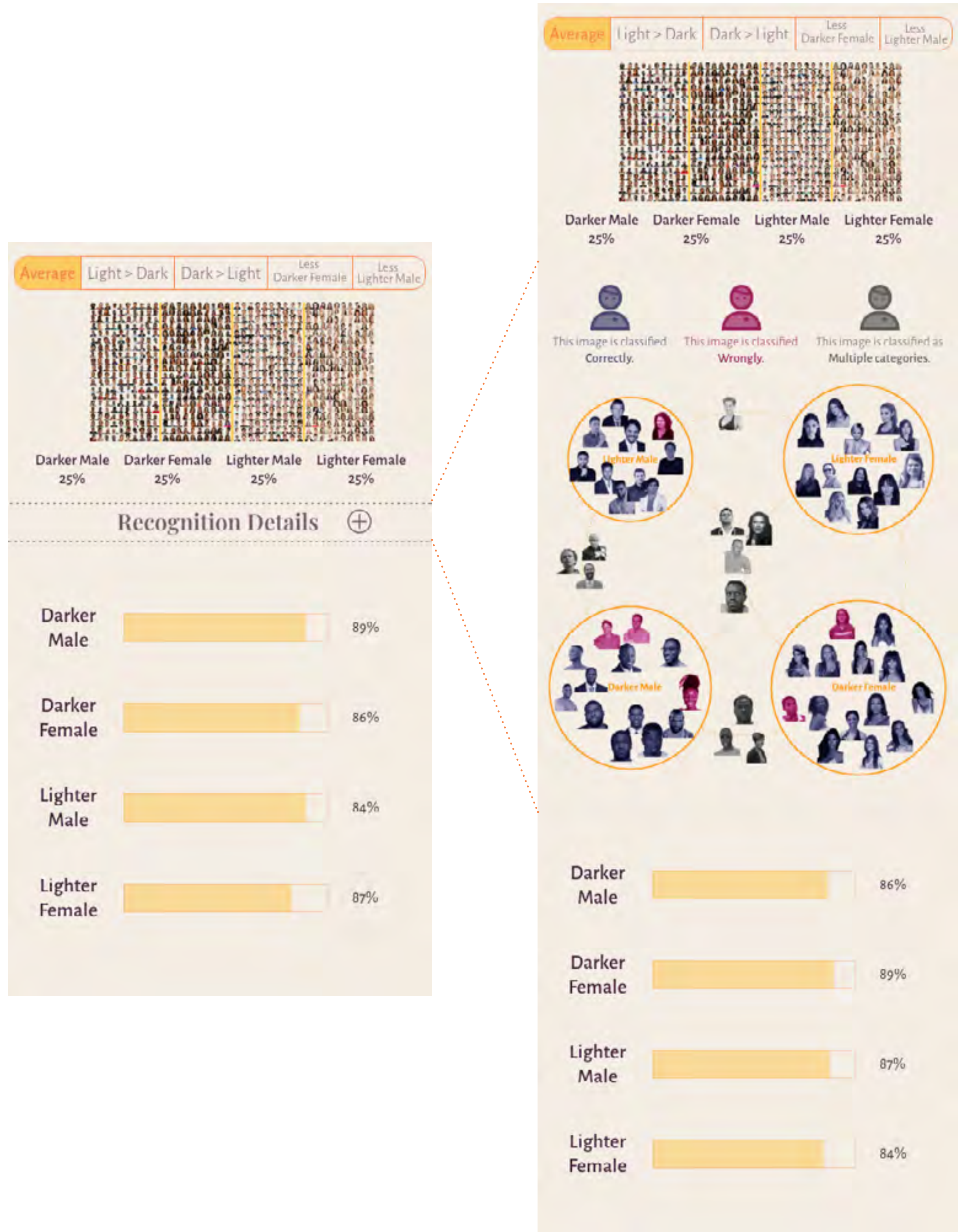


Figure. 61 Interface of the second visualisation

EPILOGUE

4.1 Difficulties during research

From the perspective of communication design students, this project has a massive amount of information from literature, from online media, about computer science, about data visualisation, and so on. The main challenge comes from how to sift through the information and use it appropriately: to strengthen the rationality of the research, to inspire the design of the output phase, and possible expressions that can be used for reference. Looking back now, I could have been more decisive and precise in the previous research, and only by decisively discarding useless information can I get valuable insights faster.

From the perspective of interdisciplinary research students, my foundation of mathematics is relatively weak, and it is difficult to read professional articles. At the beginning of the research, I positioned myself as a lay user who is just getting started. Still, as the reading literature increases, I have gradually changed to the mid-level of the audience, Machine Learning beginner, but I didn't realise it at all. This resulted in the design of the visualisation works being too technical to be understood by average users. After being aware of this problem in communication with the professors, the focus was on how to explain the complex process of image recognition by ConvNet in a straightforward way. There is another problem that I find very interesting. In the middle and late stages of the research, at the time when I should have been designing, I still did not stop exploring the direction of computer science. The main reason for reflection is that the knowledge I have is not comprehensive enough, for fear that the wrong information will mislead the audience. The starting point is good, but since the audience is also the general public, there is no need to dig deeper. The extent of how much we should learn from other disciplines seems to be dynamic, and appropriate adjustments should be made according to the actual situation of different projects.

4.2 Conclusion

The biggest gain of this research is that I have gained a lot of in-depth understanding of Machine Learning and other technologies. It is not as literally explained that machines have learned to learn. Humans have trained them through codes and sample files. Therefore, the so-called algorithmic bias is a social stereotype that has been amplified by computer programs. If users can comprehend this information by browsing my webpage, I can confidently declare that this is meaningful research.

From the perspective of information visualisation, the author obtained the data of visualisation works through fieldwork-like experimental methods in this research. In the experiment, the output of the algorithm is observed by fine-tuning the input variables. In this way, the individual properties and characteristics of ConvNet whenrecognisingg images can be spied on.

From the perspective of web design, the author used some knowledge of previous user experience design in the process of designing the mobile interface. By allowing users to operate some steps to establish the top level is concise content, the bottom level is a more complex content Information structure. While ensuring a lightweight reading experience, it can also objectively state the complex professional range of the neural network.

REFERENCE

Bhandare, A., Bhide, M., Gokhale, P., & Chandavarkar, R. (2016). Applications of convolutional neural networks. International Journal of Computer Science and Information Technologies, 7(5), 2206-2215.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91).

Choo, J., & Liu, S. (2018). Visual analytics for explainable deep learning. IEEE computer graphics and applications, 38(4), 84-92.

Chung, S., Park, C., Suh, S., Kang, K., Choo, J., & Kwon, B. C. (2016). Re-VACNN: Steering convolutional neural network via real-time visual analytics. In Future of interactive learning machines workshop at the 30th annual conference on neural information processing systems (NIPS).

Dealing With Bias in Artificial Intelligence
https://www.nytimes.com/2019/11/19/technology/artificial-intelligence-bias.html

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. Biological cybernetics, 20(3-4), 121-136.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.

Google Colab Free GPU Tutorial.
https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d

Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software
https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/#5f16b2cc713d

Harley, A. W. (2015, December). An interactive node-link visualization of convolutional neural networks. In International Symposium on Visual Computing (pp. 867-877). Springer, Cham.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Hill, C. A., & Helmers, M. (Eds.). (2012). Defining visual rhetorics. Routledge.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504-507.

Hohman, F., Kahng, M., Pienta, R., & Chau, D. H. (2018). Visual analytics in deep learning: An interrogative survey for the next frontiers. IEEE transactions on visualisation and computer graphics, 25(8), 2674-2693.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology, 160(1), 106.

Image Classification With Convolutional Neural Network | by Black_Raven (James Ng)
https://towardsdatascience.com/using-artificial-neural-network-for-image-classification-9df3c34577dd

Image Recognition: Dogs vs Cats! (92%)
https://thedatafrog.com/en/articles/dogs-vs-cats/

ImageNet Large Scale Visual Recognition Competition (ILSVRC)
http://image-net.org/challenges/LSVRC/

Kahng, M., Andrews, P. Y., Kalro, A., & Chau, D. H. P. (2017). A cti v is: Visual exploration of industry-scale deep neural network models. IEEE transactions on visualization and computer graphics, 24(1), 88-97.

Kahng, M., Thorat, N., Chau, D. H. P., Viégas, F. B., & Wattenberg, M. (2018). Gan lab: Understanding complex deep generative models using interactive visual experimentation. IEEE transactions on visualization and computer graphics, 25(1), 1-11.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), 1995.

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. In Advances in neural information processing systems (pp. 396-404).

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Li, M., Zhao, Z., & Scheidegger, C. (2020). Visualizing Neural Networks with the Grand Tour. Distill, 5(3), e25.

Liu, M., Shi, J., Cao, K., Zhu, J., & Liu, S. (2017). Analyzing the training processes of deep generative models. IEEE transactions on visualization and computer graphics, 24(1), 77-87.

Liu, M., Shi, J., Li, Z., Li, C., Zhu, J., & Liu, S. (2016). Towards better analysis of deep convolutional neural networks. IEEE transactions on visualization and computer graphics, 23(1), 91-100.

Manovich, L. (2011). What is visualisation?. Visual Studies, 26(1), 36-49.

Nielsen, M. A. (2015). Neural networks and deep learning (Vol. 2018). San Francisco, CA: Determination press.

Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., & Mordvintsev, A. (2018). The building blocks of interpretability. Distill, 3(3), e10.

Pezzotti, N., Höllt, T., Van Gemert, J., Lelieveldt, B. P., Eisemann, E., & Vilanova, A. (2017). Deepeyes: Progressive visual analytics for designing deep neural networks. IEEE transactions on visualization and computer graphics, 24(1), 98-108.

Rauber, P. E., Fadel, S. G., Falcao, A. X., & Telea, A. C. (2016). Visualizing the hidden activity of artificial neural networks. IEEE transactions on visualization and computer graphics, 23(1), 101-110.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). Learning internal representations by error propagation (No. ICS-8506). California Univ San Diego La Jolla Inst for Cognitive Science.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), 211-252.

Schoffelen, J., Claes, S., Huybrechts, L., Martens, S., Chua, A., & Moere, A. V. (2015). Visualising things. Perspectives on how to make things public through visualisation. CoDesign, 11(3-4), 179-192.

Shankar, S., Halpern, Y., Breck, E., Atwood, J., Wilson, J., & Sculley, D. (2017). No classification without representation: Assessing geodiversity issues in open data sets for the developing world. arXiv preprint arXiv:1711.08536.

Show me - New ways of visualising data https://www.economist.com/special-report/2010/02/27/show-me

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484-489.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484-489.

Smilkov, D., Carter, S., Sculley, D., Viégas, F. B., & Wattenberg, M. (2017). Direct-manipulation visualization of deep networks. arXiv preprint arXiv:1708.03788.

Smilkov, D., Thorat, N., Nicholson, C., Reif, E., Viégas, F. B., & Wattenberg, M. (2016). Embedding projector: Interactive visualization and interpretation of embeddings. arXiv preprint arXiv:1611.05469.

Strobelt, H., Gehrmann, S., Pfister, H., & Rush, A. M. (2017). Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. IEEE transactions on visualization and computer graphics, 24(1), 667-676.

Sun, Y., Wang, X., & Tang, X. (2015). Deeply learned face representations are sparse, selective, and robust in Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2892-2900).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deep face: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

Valueva, M. V., Nagornov, N. N., Lyakhov, P. A., Valuev, G. V., & Chervyakov, N. I. (2020). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Mathematics and Computers in Simulation.

Wang, J., Gou, L., Yang, H., & Shen, H. W. (2018). Ganviz: A visual analytics approach to understand the adversarial game. IEEE transactions on visualization and computer graphics, 24(6), 1905-1917.

Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., ... & Chau, D. H. (2020). CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. arXiv preprint arXiv:2004.15004.

When It Comes to Gorillas, Google Photos Remains Blind
https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

Wikipedia - David H. Hubel
https://en.wikipedia.org/wiki/David_H._Hubel

Wikipedia – ImageNet
https://en.wikipedia.org/wiki/ImageNet

Wikipedia - Neocognitron
https://en.wikipedia.org/wiki/Neocognitron

Wongsuphasawat, K., Smilkov, D., Wexler, J., Wilson, J., Mane, D., Fritz, D., ... & Wattenberg, M. (2017). Visualizing dataflow graphs of deep learning models in tensorflow. IEEE transactions on visualization and computer graphics, 24(1), 1-12.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015). Understanding neural networks through deep visualization. arXiv preprint arXiv:1506.06579.

Yu, R., & Shi, L. (2018). A user-based taxonomy for deep learning visualization. Visual Informatics, 2(3), 147-154.

Zeiler, M. D., & Fergus, R. (2014, September). Visualising and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2019). Dive into deep learning. Unpublished Draft. Retrieved 19, 2019.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. W. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. arXiv preprint arXiv:1707.09457.

从性别歧视到招聘不公，如何让 AI 公平待人？
https://www.36kr.com/p/1725393616897

卷积神经网络 CNN 完全指南终极版（一）
https://zhuanlan.zhihu.com/p/27908027

卷积神经网络 CNN 完全指南终极版（二）
https://zhuanlan.zhihu.com/p/28173972

人脸识别有歧视？识别准确率黑人比白人差很多，厂商们这样回应
https://www.sohu.com/a/222265176_114877

深度学习卷积神经网络大事件一览
https://blog.csdn.net/u013088062/article/details/51118744

ACKNOWLEDGE-MENTS

I would like to express my gratitude to all those who helped me during the writing of this thesis.

I gratefully acknowledge the help of my supervisor, Professor Mauri and co-supervisor PhD candidate Beatrice Gobbo, for their valuable suggestions of my graduation thesis in the past nine months, which greatly improved my understanding of academic writing and taught me a lot of specific research skills. Without her patient instruction, insightful criticism and expert guidance, this thesis could not have reached its present form

Then my thanks would go to my beloved family for their loving considerations and great confidence in me all through these years. I also owe my sincere gratitude to my friends and my group members who gave me their help and love in listening to me and helping me work out my problems during the difficult quarantine days.