

ARTIFICIAL INTELLIGENCE AND COMPUTER VISION – A MATCH MADE IN HEAVEN?

(Date received: 03.12.2020/Date accepted: 13.06.2021)

Mei Yuan Koay¹, Selina X.C. Loh¹, Xiu Kai Lim¹, Weng Kin. Lai¹, Kam Meng Goh¹,
Kar Wey Leong², Tomas Maul², Iman Yi Liao², Eric Savero Hermawan²

¹Tunku Abdul Rahman University College, Kuala Lumpur

²Nottingham University Malaysia Campus

Corresponding author: laiwk@tarc.edu.my

ABSTRACT

After becoming independent in 1957, Malaysia continued as an agricultural country but quickly grew into a manufacturing nation in a relatively short time. Literally from nowhere, the manufacturing sector now commands more than 38% of the nation's GDP overtaking the agriculture sector which commands just slightly above 7%. In addition to the multinational manufacturers who are mainly in the electrical and electronics sectors, there are also other smaller producers who produce for the rest of the world. Nevertheless in order to compete, they cannot just rely on manual labour whether local or foreign, to produce high volume and high quality goods at a competitive price. With intense competition, even the old way of making many products to satisfy the global appetite for good products from both the brick-and-mortar shops to your huge online shops is no longer adequate. Manual operations in the manufacturing process can come in various forms, ranging from the very simple but monotonous and repetitive to the highly complex or sophisticated. In the quality department many of the local manufacturers have chosen to use human labour to ensure their quality is maintained. For many of these highly repetitive but relatively simple tasks, the human operators need to be properly trained for an appropriate length of time before they can perform effectively. Other than the intelligence of these operators, their ability to detect deviations from the desired patterns are also utilised. And this is where artificial intelligence and computer vision can help. This paper shows how artificial intelligence combined with computer vision can be used to improve productivity and effectiveness in three different areas within a typical supply chain to make each of this production processes more efficient, namely in the production of edible bird's nest, retail ecommerce and finally the recognition of cancerous cells.

Keywords: artificial intelligence, computer vision, image processing, automation, quality control, healthcare, ecommerce

1.0 INTRODUCTION

Malaysia's humid tropical climate and seasonal monsoonal weather have produced conducive conditions for plant growth. Hence for a long time after independence, agriculture forms the backbone of the Malaysian economy where the agriculture sector contributed 37% of Malaysia's GDP [1]. In those early years, the economy was supported by a range of primary products (e.g. tin, rubber, palm oil, timber, oil, liquefied natural gas, etc.) before the Government embarked on an industrialisation programme to transform it into a manufacturing-based economy, with electronics and electronic components as the mainstay [2]. The manufacturing sector's contribution stood at 38.3%, over a third of the country's GDP in 2018 but by now the contribution by the agriculture sector is just slightly over 7.3% [3]. Manufacturing fulfils an important need where handcrafted or manual production cannot handle, i.e. more consistent quality, cost reduction and overall reduction in production time.

Compared with mass production, manual production does not produce goods with consistent quality. This is especially true when we use human operators to do quality checks as both boredom and fatigue can cause lapses in concentration leading to

defects escaping detection and inadvertently sent to the customers. While there is demand for hand-crafted products, however if the demand is high this mode of production is not efficient. However depending on the complexity of the product manufactured as well as other factors, the manufacturing process may not be fully mechanicalised. Moreover, there are still many of the repetitive processes in the Malaysian manufacturing sector that are still handled by humans.

Because of today's very short time-to-market deadlines, demanding customers, and a rise in the complexity of products, manufacturers are finding it increasingly harder to maintain high levels of quality. And this is where artificial intelligence and computer vision can help. Prior to 1955 when there was no such term, intelligent systems were either known as thinking machines, cybernetics, automata theory, or complex information processing [4]. The term artificial intelligence was first coined at the *Dartmouth Summer Research Project on Artificial Intelligence*, a 2-month, 10-man study of artificial intelligence initiated by John McCarthy in 1956. While there are many definitions, Ray Kurzweil, an American inventor and futurist defines it as machines that perform functions that

require intelligence when performed by humans [5]. A well known global media company did a survey on the importance of artificial intelligence (AI) in the automotive and manufacturing sectors. They found that while 44% classified AI as “highly important” in improving the manufacturing function in the near term, just lightly less than half of them believed that this is “absolutely critical to success.” [6]. Hence this reaffirms that artificial intelligence holds the key to future growth and success, especially for those industries in the manufacturing sector.

Computer vision (CV) as the name suggests, fundamentally deals with how computers can gain high-level understanding from digital images or videos [7, 8]. It seeks to understand and imitate the capability of the human visual system which includes the acquisition, processing, analyzing and understanding digital images to help make better decisions. This is illustrated in Figure 1.

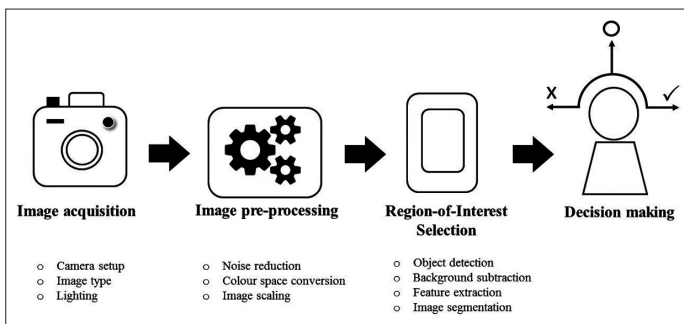


Figure 1: Traditional computer vision pipeline

This is especially true when it comes to quality control and inspection. Even the most eagle-eyed human inspector would fail at finding defects that are either very small or can only be seen at a certain angle. However a machine equipped with a powerful camera that is many times more sensitive than the naked eye will not fail. It starts with getting the human expert to identify the defects and then train the machine to perform the same task based on the results from the human expert. However, for many of these manual quality inspection, it is neither easy nor simple as it requires a huge amount of resources to properly train the humans before they can handle the task effectively.

A major driver for the industry to adopt these new technologies has been the huge improvements in image acquisition devices making them very affordable and attractive to adopt and the same goes for computing power. In tandem with these two major developments has been the intense work into the development of powerful algorithms.

For a typical product supply chain, it usually starts with the sourcing of the relevant raw materials after which these are then combined to manufacture the product before eventually reaching the consumers or customers through the appropriate distribution channel which can now be via the internet. The supply chain would generally be the same for each industry, whether it is for the production of tee shirts or medical devices. To illustrate both the flexibility as well as the power of AI and CV working together in a particular intelligent system, we have identified three key areas within such a supply chain.

Edible bird’s nest (EBN) are nests formed from the saliva of swiftlets commonly found in South East Asian and is a multi-billion dollar industry. Trained operators are currently used to sort

these natural products into the correct grade for eventual sales. In the next section we shall look at how AI working together with CV can be used to correctly grade EBN.

Breast cancer is a group of diseases in which cells in breast tissue change and divide uncontrolled, typically resulting in a lump or mass. What makes it deadly is that there are usually no symptoms when the tumour is small and most easily treated. Regular screening is important as it can help in early detection. Breast cancer is the most common cancer affecting women, with about 2.1 million cases reported per year, as communicated by the World Health Organisation (WHO). Moreover, breast cancer is also the main cause of cancer-related deaths in women. For example, it is estimated that in 2018, 627,000 women died of breast cancer, corresponding to about 15% of cancer-related deaths in women [9]. In Malaysia, breast cancer tops amongst the various types of common cancer. When cancer is suspected, a sample of the tissue for microscopic analysis is usually obtained. However, lack of trained experts, time consuming process, difficult in identifying the tumours, etc, are some of the challenges which can slow this process down. In the second part we will show some of the results of the collaboration of AI with CV to detect breast cancer using the concept of artificial teachers to help train such an intelligent system to help the medical experts. And lastly we will illustrate some of our work in using a combination of AI and CV to perform visual fashion understanding which is of huge interest for the industry, motivated primarily by the needs of electronic commerce in clothing retrieval, fashion recommendation, etc. A fundamental driver for this is to identify fashion landmarks which are functional key-points defined on clothes, such as corners of neckline, hemline and cuff.

2.0 IDENTIFYING THE RIGHT EDIBLE BIRD’S NEST

EBN are nests formed from the saliva of swiftlets commonly found in South East Asian countries of Malaysia, Indonesia, Thailand, and Vietnam [10] and is a multi-billion dollar industry. They are especially popular among some Asian communities due to their high-protein nutrients [11] and sialic acid which is believed to improve brain function.

To cater to the various customer preferences, EBNs are sold either in the unprocessed or raw form or they can also be sold after cleaning, further processing and then graded. The price of EBNs solely depends on the grade. Among the various grades, Grade AA nests are the most expensive as it is the highest quality, followed by Grade A, B and C. The grades of EBNs are normally based on their shape, size, colour, and level of contaminants. Conventionally, trained human operators would grade EBNs based on these features. However, the consistencies of the grading results can be unreliable as the manual grading process is prone to the subjectivity of the human operators [12]. Thus, such a tedious grading process leads to production costs, subjectivity in the end results, as well as low efficiency as huge amounts of time are spent on the inspection. Furthermore, hiring and training new operators to inspect the nests are unavoidable for such a conventional approach to grading. Notwithstanding the conventional method, with the technological advancement in recent decades some aspects of identifying EBNs have improved dramatically in terms of speed, accuracy, and reliability [10].

In this section, we will describe a novel and fast multi-features neural fuzzy approach using a set of unique features to automatically grade EBN and the results compared with several state-of-the-art AI (artificial intelligence) techniques. Unlike other popular approach our contributions are as follows:

- 1) a novel application of a neural fuzzy approach, based on an intuitive multi-feature set to accurately grade EBNs,
- 2) unlike other more popular AI techniques, our proposed approach does not need huge amounts of training data - which would require huge amounts of human effort and time to compile,
- 3) training is fast and it does not need special high performance computing platforms.

Grading of the EBN that is practised by the industry is based on the size, amount of contaminants, and the overall shape. A common method adopted by the industry to assess the size is by comparing it with the three fingers of the human hand where the three fingers are the index finger, middle finger and the ring finger of an average adult [13]. Clearly this can be very subjective when the operators are trained to assign the correct grade to each specimen as the size of their fingers would vary. A more scientific approach would be to measure the width and length of each specimen of the EBN but obviously this would be very time consuming.

Moreover, the grade of the EBN also heavily relies on the amount of contaminants in each sample. These contaminants or impurities can range from feathers, sand particles, bird droppings, etc. EBNs with lesser amount of contaminants will fetch a higher price compared to those with more contaminants. Finally, the shape of the harvested nests will have a significant effect on the grade. Deformed EBNs that are significantly different from their accepted shapes will have a lower commercial value [13]. What this means is that specimens when placed evenly on a horizontal surface that exhibit a perfect half cup shape would be a good grade. Lower grades, would also have a similar shape as their better counterparts but these will not be even when placed on a horizontal surface, as shown in Figure 2.

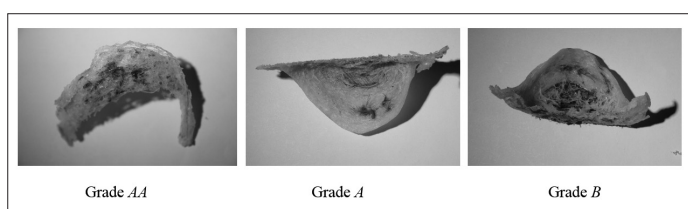


Figure 2 Examples of EBN for Grade AA, Grade A and Grade B. Adapted from [Introduction to Birds Nest 2018]

An important element for any successful autograding system is CV. For example, to identify the correct grade of mangoes is to correlate it to the fruit's maturity level. In such an intelligent system when used to auto grade and sort agricultural produce like mangoes, CV would again play a very important role as a core component. Nandi *et al.* first obtained sample images of unsorted mangoes for five different varieties of mangoes. Human experts were then used to predict the maturity of mangoes from these images. In their study, the Gaussian Mixture Model (GMM) was used to predict the maturity level of the mangoes based on the set of relevant features of the mangoes. They showed that the classification accuracies of the Gaussian Mixture Model (GMM)

in their system were as good as those of human experts [14].

Diseases of fruits like apples, palm oil fruits, mangoes, dates and strawberries can also be correctly identified with the appropriate classifier working on the extracted features of the fruit. In order to correctly identify the disease, a unique set of features were used for each different fruit. Solanki *et al.* applied such an approach involving the right classification technique with an effective set of extracted features to identify the diseases [15]. They segmented the images of the fruits in order to identify and segregate the healthy and infected parts on the fruit. With these results, the system was then able to correctly grade the fruits based on the amount of infections found on the fruits. To improve the system's performance various segmentation techniques to identify the region of interest were also investigated and these include histogram matching, region based segmentation, edge detection and fuzzy segmentation.

Automated classification and grading of agricultural products have been extensively implemented using a variety of different classification techniques [16, 17, 18, 19, 20, 21]. Other than agricultural produce, Han Yan *et al.* developed an Adaptive Neuro-Fuzzy Inference System (ANFIS) to classify water quality status of river based on a total of 9 weeks data (845 observations consisting of 3 water quality parameter) collected from 100 monitoring stations in all major river basins in China [22]. They were able to report good classification accuracies with such an approach. Other than water quality identification, B. Ari Kuncoro and Suharjito investigated ANFIS to classify the texture of crumpled aluminum foil, corduroy, cotton and orange peel from the images using a combination of Discrete Cosine and Gray Level Occurrence Matrix (GLCM) methods on the set of extracted features [23]. Other areas where ANFIS was used includes the work by Alireza *et al.* where ANFIS is now used to detect and diagnose breast cancer based on a set of risk factors [24]. In their work, they chose to use ANFIS to see if it can correctly identify breast cancer patients using nothing more than a set of 22 features from the standard data sets and then on real data. They were able to obtain satisfactory results with the ANFIS they developed.

2.1 Auto Grading of EBN

The EBN images were collected from a centralised EBN processing facility with modern and hygienic facilities. This is a more cost-effective way of processing the EBN. The EBN when harvested is sent for processing by the individual EBN farms which are usually managed on a small scale. They would not be able to provide either the facilities or the human personnel to process the raw EBN and grade the EBN efficiently and consistently.

To test the auto grader, we used the EBN of 3 different grades, AA, A and B which had been carefully graded by the trained operators. A total of 63 samples were collected and photographs were taken with a standard Digital SLR camera fitted with a macro-lens.

Table 1 shows a selection of the 4 extracted features for each class of EBN. Notice the different range of values for each feature. Both the size and the estimate of the impurities are in their thousands whereas curvature and colour are less than one.

Classification of the EBN grades has been investigated with several powerful and popular AI techniques, namely the

Adaptive Neuro Fuzzy Inference System (ANFIS), k-Nearest Neighbour (kNN), Naive Bayesian (NB), Support Vector Machines (SVM), as well as Artificial Neural Networks (ANN). Because the values of the extracted features are of different ranges, we need to normalise them to the range [0, 1] to avoid the larger numerical values from any of the features dominating the auto-grading process. Hence we investigated several different data preprocessing methods that converted the raw data into an effective and efficient form before the data were processed. Z-score normalization, decimal scaling, linear normalization, min-max normalization and a non-linear normalization based on the sigmoid function were investigated. The results showed that the classifiers are sensitive to the pre-processing schemes used and they performed better for some of these normalisation schemes. kNN was not able to match the accuracies of the others managing only 61.54%. On the other hand, NB was able to match the accuracy achieved of the more modern SVM classifier. In addition, if we just look at the lowest performance of ANFIS, it still outperforms most of the other classifiers as shown in Table 2.

Table 1: Features extracted for various grades

grade	curvature	size	contiminants	colour
AA	0.3560	5784956	1148060	0.6824
AA	0.7969	2646064	2370178	0.6784
-	-	-	-	-
A	0.2403	6333746	133551	0.7608
A	0.4217	5075468	416805	0.7765
-	-	-	-	-
B	0.3537	5361693	518880	0.6784
B	0.7216	5084135	4971157	0.6980
-	-	-	-	-
-	-	-	-	-

Table 2: Maximum and minimum classification accuracies (%)

Classifier	Min	Max
kNN	54.19	61.54
NB	72.09	89.70
ANN	68.32	89.28
SVM	74.97	89.52
ANFIS	81.34	94.09

Finally, the Convolutional Neural Network (CNN) was also tested to auto-grade EBN. The MVGG-19 [25] which is based on the VGG-19 CNN and uses the same number of convolutional layers and max pooling layers was chosen.

The MVGG-19 used here has 19 layers, with 32 input filters in the first layer. However, the MVGG-19 has a smaller number of parameters compared to the VGG-19. The MVGG-19 was integrated with Keras library and TensorFlow backend. Moreover as it is computationally intensive, the NVIDIA GeForce GTX 1070 GPU was used to speed up the computation. Since the EBN dataset images have large dimensions, the computation time is much longer. Hence, the dataset images were scaled down before they can be processed by the CNN model. Squashing was adopted to resize the images. Furthermore, as such computation methods thrive on large data sets, data augmentation in the form of random transformations were applied to the existing EBN

dataset such as horizontal/vertical flipping, clockwise and anti-clockwise rotation. Table 3 summarises the best classification accuracy extracted from the various pre- processing schemes.

Table 3: Best Classification accuracies obtained

Classifier	Accuracy
kNN	61.54
NB	89.70
ANN	89.28
SVM	89.67
ANFIS	94.09
MVGG-19	88.89

With 4 features presented to the classifiers, ANFIS performed better than kNN with the kNN able to only achieve 62% whereas ANFIS was able to achieve 94%. Neural nets, SVM and NB were able to achieve about 89%, some 5% less accurate than ANFIS. We have also evaluated a non-feature based approach that does not require any feature extraction as it uses the images. However, as the data set is relatively small, a significant amount of preparatory work needs to be done before the MVGG-19 CNN was able to learn the various grades and then autograde the remaining images not used in the training. The CNN was able to produce an average accuracy of 89% with this non-features approach.

A major advantage of our features-based approach is that it is less time consuming as the system developed here was able to achieve satisfactory results without requiring a huge set of data nor any additional computing hardware. In the future, we would like to extend the work to look at EBNs from a different source where the grading may be different as well as to investigate a better imaging system to further improve the accuracy of the system.

3.0 ARTIFICIAL TEACHERS FOR BREAST CANCER CLASSIFICATION

The field of Artificial Intelligence is particularly compatible with the overall approach adopted by the field of Natural Computation, where researchers aim to solve computational problems by getting inspiration or adapting mechanisms from nature [26]. One of the earliest examples, pioneered by McCulloch and Pitts [27], corresponds to a family of techniques, which form the foundation of the current AI explosion, i.e. deep learning, also known as artificial neural networks, more generally. The process of obtaining inspiration or adapting mechanisms from nature is however, far from complete. In fact, it could be argued that it has barely begun. Other examples of neural network breakthroughs associated with bioinspiration include convolutional neural networks [28], and the usage of attention-based mechanisms [29], all of which have been particularly prominent in targeting CV problems.

When seeking bioinspiration for Artificial Intelligence, it is natural to focus on the biological organism that exhibits the highest level of general intelligence so far, i.e. humans. When we compare arguably the most common learning paradigm in the field of neural networks (i.e. supervised learning with a single dataset) with how teaching/learning occurs in humans, the constraints and limitations of the former become very evident. For example, humans learn a diversity of different fields and skills, in carefully structured stages, with different kinds of assessments, involving

a variety of teachers and teaching methods. In other words, the simplicity of neural network learning paradigms is currently no match for the richness and complexity of the educational approaches and pedagogical techniques practised by humans. Having said this, researchers have gradually been making some progress in this direction. For example, in curriculum learning Bengio *et al.* [30] divide the learning process into several stages of increasing complexity, which partially addresses the approach of learning in “carefully structured stages” mentioned above. Furthermore, and in direct connection with this section, Hinton *et al.* [31] have improved the classification accuracy of a “student” model by employing an additional “teacher” model, which partially addresses the “variety of teachers” limitation mentioned above.

The approach introduced by Hinton *et al.* building upon the work of Rich Caruana and his colleagues [32] is generally referred to as knowledge distillation. The approach hinges on the key insight that crucial knowledge regarding generalization can be transferred (distilled) from a larger teacher model (e.g. an ensemble of many models or a large single model) to a smaller student model, by making the student learn the teacher’s full set of output/label probabilities rather than just the single winning output/label. The full vector of output/label probabilities contains rich information not necessarily obtainable by the small model when trained from scratch on the same training set. For example, in the classic digit classification dataset MNIST, if the input pattern is the digit “8”, then the probabilities of the output nodes corresponding to “8”, “3” and “1”, might be 0.6, 0.2, and 0.05 respectively. These relative probabilities contain useful information about digit similarities (i.e. the similarity between 3 and 8 is higher than the similarity between 1 and 8), which in turn is useful for generalization. The information contained in the full vector of probabilities is often called “dark knowledge” because it represents the hidden/implicit knowledge learnt by an ensemble, or a large regularized model. If we relate back to the human educational domain, an intuitive but loose analogy would be to equate the traditional approach of directly training on the dataset as “learning facts”, and the knowledge distillation approach as additionally learning the attitudes, soft skills, relative emphasis, and indirect implications of the teacher. We can also see that the artificial assistants also need to “go back to school” - using the concept of teacher and student to learn the key details to be effective. In this case, the neural networks have to go back to school!

As mentioned in the introduction, breast cancer is the most common cancer affecting women, and is also the main cause of cancer-related deaths in women. In Malaysia, breast cancer is the top cancer amongst the various types of common cancer [33] as shown in Figure 3. In the period between 2012 and 2016 the following were the 10 most common cancers diagnosed in Malaysians: ovary, uterus (cervical), liver, prostate gland, leukaemia, nasopharynx, lymphoma, lung, colorectal, and breast cancer.

In recent work, we have experimented with extensions to the above knowledge distillation ideas and have applied these to the domain of cancerous breast tumour classification in mammograms [34] which serves as another example of the fruitful marriage between CV and AI. In particular, images from the Digital Database for Screening Mammography (DDSM) were used. This is a database to facilitate research into AI techniques for breast cancer screening. The database was built

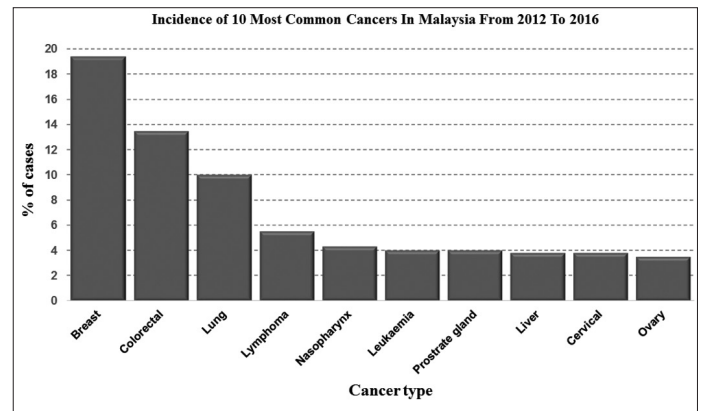


Figure 3: Incidence of 10 Most Common Cancers In Malaysia From 2012 To 2016

from about 2,500 studies, each one consisting of two images of the breast, one from the cranial-caudal (CC) view and the other from the mediolateral-oblique (MLO) view. The images are classified into the following labels: *normal*, *cancerous*, *benign* or *benign without callback*. The images are also accompanied by corresponding pixel wise segmentations of the lesions [35, 36]. To the best of our knowledge there are currently no other research works that explore knowledge distillation questions in the context of the DDSM dataset.

In particular we have sought to extend the knowledge distillation (KD) approach by looking into two key ideas: (1) the introduction of assistants between the teacher and student, and (2) reversing the direction of distillation. The first idea involves distilling from the teacher to one or more intermediate assistants before distilling from the final assistant to the student. This idea was initially proposed by Mirzadeh *et al.* [37], based on the observation that a student’s performance deteriorates as the gap between the teacher and the student increases. Motivated by this observation, the authors recommended to place assistant models in between the teacher and student to more effectively transfer the required information. The authors showed that student model performance is positively correlated with the number of assistant models used. Regarding the second idea, distillation traditionally occurs from the more complex (larger) model to the simpler one, so a reversal here implies distilling knowledge from a smaller model and applying this to a larger model. The idea of KD reversal is somewhat related to the idea of curriculum learning [30], except that instead of implementing gradually increasing complexity at the level of datasets, this is implemented at the level of models, whereby the output probabilities of simpler models are used to help the representation learning of larger models.

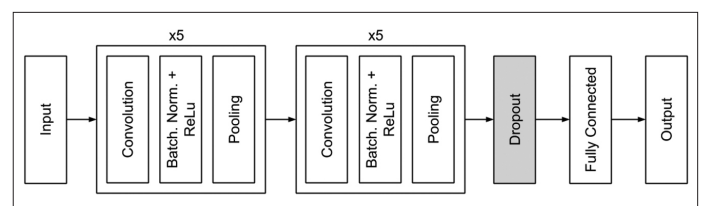


Figure 4: Our baseline model consisted of convolutional, batch normalization, pooling, dropout and fully connected layers

In our main experiment, we set up a series of conditions to test the performance of different numbers of assistants (i.e. 0, 1, or 2), and different distillation directions (i.e. the standard

flow from a more complex teacher to a simpler student, or the reverse). Our baseline model, depicted in Figure 4, consisted of ten 2-dimensional convolutional layers, each with batch normalization, ReLU activation functions and max pooling layers, and one final fully-connected layer.

All of the remaining models consisted of standard residual networks [38], adapted from the standard ResNet models provided by PyTorch in their torchvision package, adopting varying numbers of layers (i.e. 18, 34 and 50), applying dropout, and limiting the number of outputs to 3. For the performance metric reported here, we adopt the area under the curve (AUC) metric, computed relative to the malignant class. The AUC performance values for models trained in the traditional manner (without KD) for the baseline, ResNet18, ResNet34, and ResNet50 models, were 0.7684, 0.8792, 0.8810, and 0.8876 respectively, and can be found in the “Base AUC” column.

Table 4: AUC of class 2 (malignant) against class 0 (normal) and 1 (benign) for models with knowledge distillation

KD Flow	Teacher	Assistant 1	Assistant 2	Student	Base AUC	KD AUC
D→S	ResNet50			Baseline	0.7684	↑ 0.7706
S→D	Baseline			ResNet50	0.8876	↓ 0.8871
D→I→S	ResNet50			ResNet34	0.8810	↑ 0.8818
D→I→S	ResNet50	ResNet34		Baseline	0.7684	↑ 0.7726
S→I→D	Baseline			ResNet34	0.8810	↑ 0.8850
S→I→D	Baseline	ResNet34		ResNet50	0.8876	↑ 0.8893
S→I1→I2→D	Baseline			ResNet18	0.8792	↑ 0.8810
S→I1→I2→D	Baseline	ResNet18		ResNet34	0.8810	↑ 0.8891
S→I1→I2→D	Baseline	ResNet18	ResNet34	ResNet50	0.8876	↑ 0.8902

Legend: D = deep network, S = shallow network, I = intermediate-sized network compared to D and S, I1 = smaller network than I2. The knowledge is distilled from the Teacher to Assistant 1, then to Assistant 2 and lastly to the Student. The “Base AUC” column refers to the baseline performance of student models without KD (in the “Student” column) and is portrayed for the sake of easy comparison with corresponding KD variants. Arrows in the rightmost column denote whether the KD conditions improved the AUC of corresponding non-KD (base) conditions. All cases showed an improvement except for the reverse non-assistant condition (i.e. S→D).

Changing the conventional knowledge distillation method of ResNET50 to a shallower network seems to be able to boost the performance of the student based as shown here in Table 4. The reverse knowledge distillation method (i.e. S→D case) however did not improve the performance of the deeper student model. However, when one assistant was added, both KD flows (i.e. D→I→S and S→I→D) showed an increase in AUC relative to their non-KD conditions. The usefulness of assistants and reverse KD was also observed when two assistants were used, as seen in the S→I1→I2→D condition, which exhibited the best performance across all conditions (i.e. 0.8902 AUC), indicated by bold typeface in the table. Overall these results clearly show the positive effect and potential of combining assistants with reverse KD. Both techniques are motivated by the need to decrease the complexity/knowledge gap between models, and presumably both are effective for the same reason, although a detailed and insightful explanation into the effectiveness of the techniques is an open question for future work.

The techniques discussed here clearly highlight that even within the relatively narrow scope of extending learning

paradigms with multiple artificial teachers and teaching flows, the bioinspired approach to AI in general, and AI for Computer Vision in particular, still has a lot to offer.

4.0 IMPROVING EFFECTIVENESS OF ECOMMERCE

In getting an overview on how big ecommerce is we turned to a report by a global business data platform which conducted an extensive study in 2019 focusing on both sales and revenue of ecommerce. They concluded that retail ecommerce sales stands at USD3.53 trillion in 2019 but growth of revenue from e-retail would nearly double that number by 2022[39]. Online shopping would remain as one of the most popular online activities worldwide.

Among those, online clothing shopping has seen immense growth in popularity in recent years. According to Rusell [40], fashion retailers have had a 38% increase in online orders even during the Covid-19 pandemic.

However, the issue with online clothing shopping is the inconsistent sizing across different brands and manufacturers [41]. Various studies have shown that 50% of returns are caused by sizing issues [42, 43]. Furthermore, fit of the garments may vary as body proportions do not conform entirely to standard sizing. A Kurt Salmon Associates study reported that 50% of women have difficulty finding apparel that fits. To make a custom-fit garment, customers would have to manually provide their measurements, which may not be readily available or accurate. Custom-fit garments are also usually costly and inconvenient to purchase due to the need for multiple fittings to achieve the desired fit. This is where the need for virtual tailoring arises.

Deep learning techniques have been widely applied to industries where a large amount of labelled data can be collected/acquired, e.g., in wireless communication [44], industrial robotics [45], smart manufacturing [46], and quality control in printing industry [47], to name a few. However, data labelling process can be tedious, time-consuming, and prone to human errors. This is particularly so in the case of labelling accurate landmark positions on object/subject of interest for a large number of images [48, 49, 50, 51, 52]. It thus poses us a critical question, i.e., is it still possible to address industrial problems where the labelled data sample size is small?

The literature suggests that such problems may be addressed by appropriately transfer-learning from existing Deep Learning models on similar problems and with data augmentation. A recent study on smart buildings has shown that a deep neural network trained with building source data in Beijing has been successfully adopted and re-trained to predict ventilation control for buildings in Shanghai with just 15 days’ of data [53]. On the other hand, for data such as images that have much higher dimensionality than that of a smart building, data augmentation has been playing significant role for boosting up small samples [54].

In this section, we report a case study on body landmark detection with an extremely small dataset. Figure 5(a) shows a typical example of the frontal image of a customer and the expected body landmarks, including five landmarks along the shoulders (‘Shoulder.A’ to ‘Shoulder.E’), two landmarks on the right and left side of the chest (‘L.Chest’ and ‘R.Chest’), two landmarks on the right upper arm (‘Arm.A’ and ‘Arm.B’), and finally one landmark on the left wrist (‘Arm.E’). To the best of our knowledge, there is no public dataset that is similar to the

body landmark dataset shown in Figure 5. As there are currently only 99 sample images with their corresponding landmarks, it is far from being adequate to train any deep neural network from scratch.

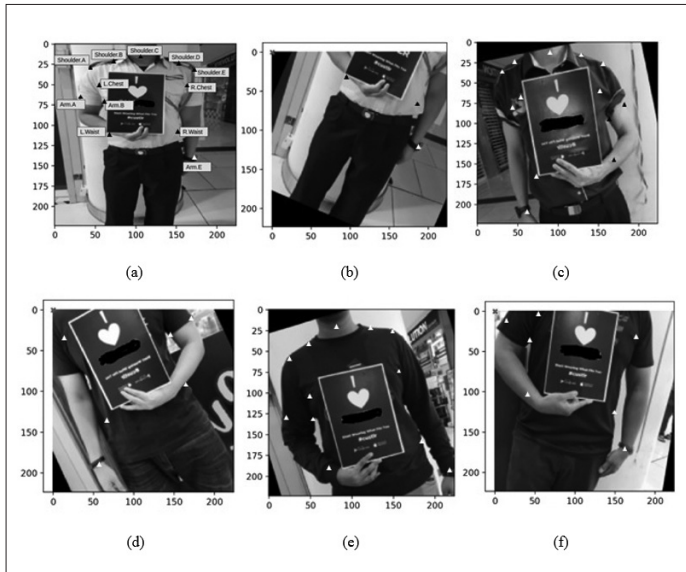


Figure 5: Examples of augmented training images with ground truth body landmarks (triangles). (a) Labels of body landmarks. (b)-(f) augmentations through rotation, translation, cropping, flipping, etc. of different training images and their corresponding ground-truth body landmarks.

However, there do exist public datasets in related domains, e.g., DeepFashion [55], Pose [56, 57], etc. Therefore, by leveraging on the concept of transfer learning, we explore architectural modification to an existing Deep Fashion Net that was trained for fashion landmark detection [58], and re-train the modified deep convolutional neural network for body landmark detection. We divided 99 samples into 79 training samples and 20 validation samples. Due to the extremely small sample size, we applied heavy online augmentation where each sample is augmented 100 times with random rotation, translation, cropping, and left-right, leading to a total sum of 7900 training samples and 2000 validation samples, respectively. Figure 5(b)-(f) show some examples of augmented images and their corresponding landmarks (triangles).

The DeepFashionNet in Figure 6 contains a base network (i.e., VGG16) for extracting image features and classifying clothes to different attributes and categories, a landmark branch for predicting fashion landmarks (see the circles in Figure 8), and an attention unit that focuses the classification on the area of fashion landmarks.

The proposed Fashion-to-Body Attention Network adopts the VGG16 feature extractor, fashion landmark branch, and the attention unit in the DeepFashionNet, but replaces the classification branch with a body landmark branch (see Figure 7). The body landmark branch has identical architecture to the fashion landmark branch, except that the output heatmap has 12 channels corresponding to 12 body landmarks, instead of 8 channels corresponding to 8 fashion landmarks. The loss function is defined as the mean squared error (MSE) between the output heatmap and the ground-truth heatmap.

To train the proposed model, the weights for VGG16 extractor and fashion landmark branch are loaded from the trained

DeepFashionNet [59] and are frozen. The weights of the attention unit are initialised with uniform distribution on $[0, \text{total number of weights in the network}]$. The body landmark branch are initialised according to Kaiming He method [60]. The learning rate is set as 0.0001 with decay. The optimisation method is the Adam optimiser [61]. The training would run for 20 epochs in total.

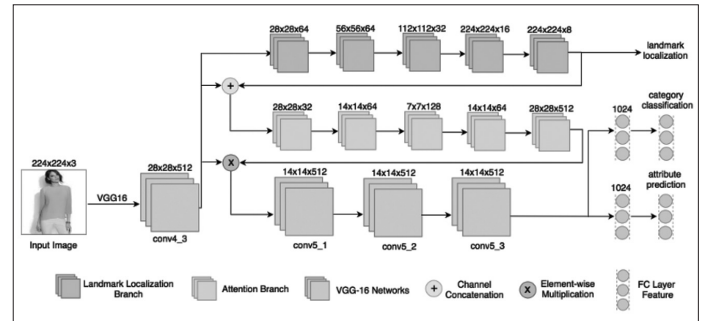


Figure 6: Deep Fashion Analysis [52]. The architecture contains a base network of VGG16, a fashion landmark branch, and an attention unit.

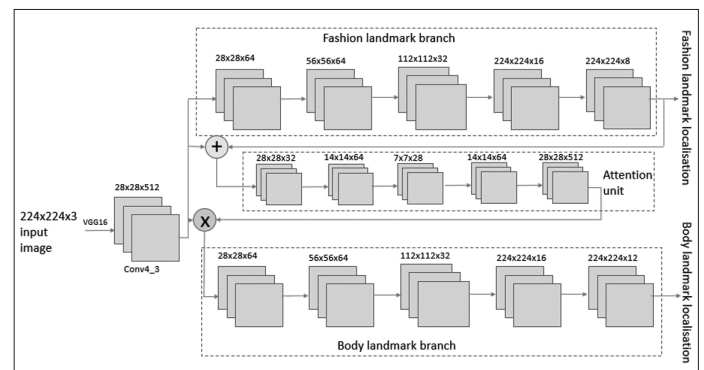


Figure 7: The proposed Fashion-to-Body Attention Network for body landmark detection. It replaces the classification branch in the base network with a body landmark detection branch. The idea is to use fashion landmarks detected by the fashion landmark branch to provide the attention for the body landmarks.

Figure 8 demonstrates the body landmark detection results of several testing images by applying our proposed model. The crosses are the detected landmarks whilst the triangles are the ground truth. In comparison, the fashion landmarks detected by the base network, DeepFashion Net, were also plotted as circles. Notice the semantic difference between the ground-truth body landmarks (i.e., triangles) and the fashion landmarks (i.e., circles). Visually it is clear that the predicted body landmarks (i.e., crosses) are much closer to the ground-truth body landmarks than to the fashion landmarks. Table 5 has shown that the testing error in terms of normalised average landmark distance between the predicted landmarks and the ground-truth, is 0.0284 for the proposed model. It is around 47% lower than that was reported for the Fashion landmarks (0.0534) by [52]. It demonstrates that the proposed model has successfully transferred fashion landmark detection to body landmark detection through the attention mechanism.

Nonetheless, the preliminary experimental results also revealed several drawbacks of the proposed model as well as some intrinsic problems with the dataset. As shown in Table 5, the mean individual body landmark localisation error varies from landmark to landmark. The largest individual landmark localisation error occurs with the landmark 'Arm.E' and

‘L.Waist’. There is a large variation of local image features around the area where the landmark ‘L.Waist’ would be marked in different images, due to the fact that customers were holding an A4 paper with their right hands, sometimes blocking the waist. On the other hand, even if the waist is not blocked, the variation of labelling a waist landmark is still larger than other

landmarks. This is reflected as the landmark localisation error for ‘R.Waist’ is the second largest (see Table 5).

The large mean localisation error for ‘Arm.E’ over all the testing images is mainly skewed by the large error of the wrongly detected landmark in image e.g., Figure 8(c). In fact, the landmark ‘Arm.E’ in Figure 8(c) was not detected but instead a

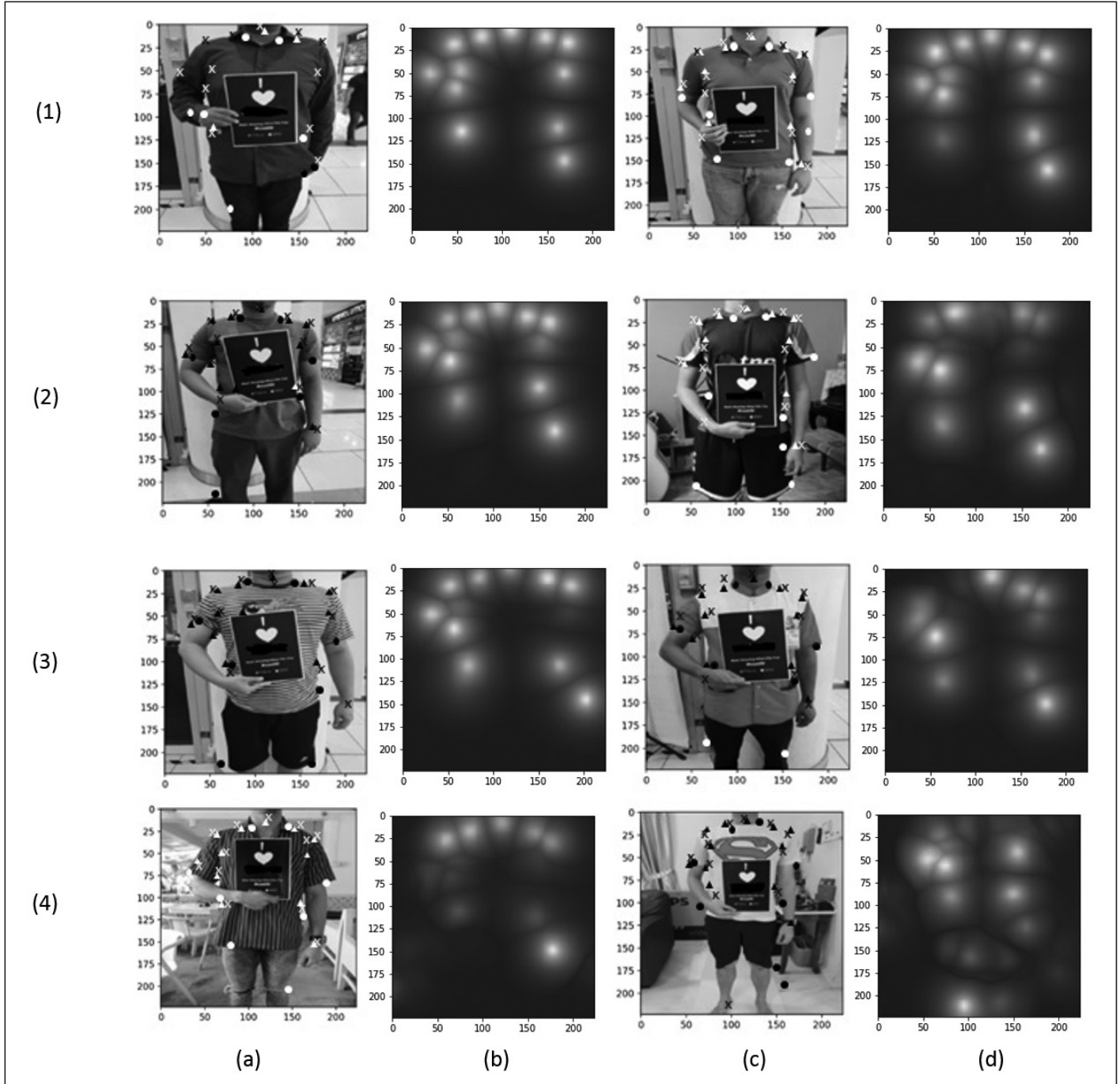


Figure 8: Columns (a) and (c) are frontal images of customers with their respective body landmarks and fashion landmarks. Triangles (Δ): ground-truth body landmarks (i.e., manually labelled); Crosses (X): body landmarks detected by the proposed model; Circles (O): fashion landmarks detected by the base network, DeepFashion net. Columns (b) and (d) are the landmark heatmaps for the images in columns (a) and (c), respectively.

Table 5: Mean body landmark localisation errors of testing images

L.Che	R.Che	Sho.A	Sho.B	Sho.C	Sho.D	Sho.E	Arm.A	Arm.B	Arm.E	L.Wai	R.Wai	Average
0.0188	0.0262	0.0221	0.0226	0.0162	0.0246	0.0135	0.0279	0.0122	0.0632	0.0634	0.0302	0.0284

point around the right ankle was detected. A close examination of its corresponding heatmap (Figure 8(d)) shows that there are more than 12 landmark responses. This is possible since the convolutional layers' receptive field is small and therefore local features that excite the relevant neurons will be detected and returned as a candidate landmark. The right ankle has similar local features as that of a wrist and thus had very high response. As the current model does not incorporate any structural constraints on the body landmarks, it returned the largest response in each channel of the output heatmap which contains the right ankle.

Observing carefully the heatmaps in Figure 8, we notice some of the detected landmarks have very weak and/or wide responses and resulted in relatively large visual localisation error compared to the rest of the landmarks, e.g., 'Shoulder.A' and 'Shoulder.B' in image (3)(c), 'R.Chest' in image 8(a). These problems may be attributed to the lack of texture variation around the landmarks in concern, i.e., the foreground looks similar to background. In such cases, it is important for the model to pick up the structural/shape constraints among the landmarks to counter the lack of texture information. This is, however, not the case with the proposed model.

We also observe that when the localisation error of individual landmarks is high, it does not necessarily mean the landmark has large semantic deviation from the ground-truth landmark, particularly in the case of 'L.Waist' and 'R.Waist'. It suggests that it is important to have a different evaluation metric that would better capture such characteristics than the current MSE error for individual landmarks.

The current dataset is small, lacks variations in poses, clothing styles, clothing texture, background, etc., and suffers from high interpersonal variation of manual landmark labelling. These issues should be addressed during the course of further data collection.

5.0 CONCLUSIONS

In this paper we have illustrated what the *marriage* of artificial intelligence which Kurzweil has conveniently compared to just a "a bag of tools" with CV can accomplish. The first section introduced how AI working together with CV can revolutionize the multi-billion dollar EBN industry by providing better efficiency and productivity. As the price of EBNs solely depends on the grade it is the current practise in the industry to use trained operators to sort the raw EBNs to the correct grade. While the human visual superior system is much better than the current cameras however, humans even trained ones can become tired easily and even inconsistent occasionally. By combining AI with CV we investigated 6 different AI techniques to autograde EBNs with ANFIS achieving an accuracy of 94.09%. Breast cancer is a major disease around the world and because it does not show any symptoms unless it is too late, early detection through regular screening is necessary for successful treatment. In the second part we showed the results of the use of CV with AI to detect breast cancer using the concept of bioinspired AI help train such an intelligent system in order to help the medical experts. Using two assistants and reverse KD, a maximum performance across all conditions of 0.8902 was obtained. And lastly we illustrated how a combination of AI and CV was used to perform visual fashion understanding which is of huge interest for the industry, motivated primarily by the needs of electronic

commerce in clothing retrieval, fashion recommendation, etc. A fundamental driver for this is to identify fashion landmarks which are functional key-points defined on clothes, such as corners of neckline, hemline and cuff. Unlike conventional AI approach on such a technique, our DeepFashionNet was trained only on an extremely small dataset, something is unconventional and challenging to do. With such an approach we were able to achieve mean body landmark localisation errors of the testing images from a low of just 0.0122 (Arm.B) to 0.0634 (L.Wai).

With AI trending at the moment and the spotlight turned on it, many in the industry wants to believe that this is the silver bullet that they have been waiting to solve all their woes. However, those working in the field will tell you that while many of the success stories published are very encouraging, there are certain limitations to AI's ability. However this is not to say that AI does not work but while progress is being made in pushing the boundaries on those which work and those which do not, new(er) applications are being announced almost on a daily basis. In this paper we have combined the "intelligence" of AI with CV to show how it can work on several different problems, as we believe the brain needs to "work with the eyes" - just like what humans normally do when we try to interpret what our eyes tell us. Many of the current approaches in using AI requires large amounts of data (EBN grading, breast cancer identification for example) where more is preferred. A major challenge will be the collection of these data as it can be challenging, time consuming and costly to assemble a large amount of data. On the bright side, with the cost of both computing power and sensors coming down it will now be more cost effective for industries to adopt such technologies. Finally we hope that the results from using this combination of AI and CV on three different areas have clearly shown they are a match made in heaven - with the potential to help produce a better world. ■

REFERENCES

- [1] Olaniyi A. O., Abdullah A.M., Ramli M. F. and Sood A. M., Agricultural land use in Malaysia: An historical overview and implications for food security, Bulgarian Journal of Agricultural Science, vol. 19 (No 1) 2013, pp 60-69.
- [2] John H. Drabble, Economic history of Malaysia, Economic History Association EH.net <https://eh.net/encyclopedia/economic-history-of-malaysia/> accessed on 18th May 2020
- [3] Statista, Malaysia: Share of economic sectors in the gross domestic product (GDP) from 2008 to 2018. <https://www.statista.com/statistics/318732/share-of-economic-sectors-in-the-gdp-in-malaysia/> accessed 20 May 2020
- [4] McCorduck, P., Machines Who Think, A.K. Peters, Ltd, Second Edition, 2004.
- [5] Kurzweil, Raymond, What is Artificial Intelligence anyway?, American Scientist Vol. 73, No. 3, 1985, pp 258 - 264.
- [6] Forbes Insight, How AI Builds A Better Manufacturing Process, July 17, 2018. <https://www.forbes.com/sites/insights-intelai/2018/07/17/how-ai-builds-a-better-manufacturing-process/#57aff681e842> accessed on 18 May 2020
- [7] Wikipedia, Computer vision https://en.wikipedia.org/wiki/Computer_vision. accessed on 20 May 2020
- [8] Technopedia, Computer vision. <https://www.techopedia.com/>

- definition/32309/computer-vision accessed on 20 May 2020
- [9] WHO - Breast Cancer <https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/> accessed on 18th May 2020
- [10] Introduction to Bird's Nest, (2018), <http://yanwo.saikim.com.my/english/introduction-birds-nest.html>. accessed on 28th March 2018.
- [11] Chua, Y. G., Chan, S. H., Bloodworth, B. C., Li, S. F. Y., and Leong, L. P., "Identification of edible bird's nest with amino acid and monosaccharide analysis" *Journal of Agricultural and Food Chemistry*, 2014, vol. 63 (1), pp. 279-289.
- [12] Goh, K.M., Lai, W.K., Ting, P. H., Koh, D., Wong, J. K. R., "Size Characterisation of Edible Bird Nest Impurities: A Preliminary Study", 21st International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES2017), pp 1072 – 1081, 6 - 8 September 2017, Marseille, France.
- [13] Tan, K. H., Chia, F. C., and Ong, A. H. K., "Impact of Swiftlet's Moulting Season on the Value of Edible Bird Nests", International Conference on Intelligent Agriculture, IPCBEE vol. 63, IACSIT Press, Singapore, DOI: 10.7763/IPCBEE, 2014, vol. 63.4.
- [14] Nandi, C. S., Tudu, B., and Koley, C., "An automated machine vision based system for fruit sorting and grading" Sixth International Conference on Sensing Technology (ICST), 2012, pp. 195–200.
- [15] Solanki, U., U. K. J. D. G. T., "A Survey on Detection of Disease and Fruit Grading", International Journal of Innovative and Emerging Research in Engineering, vol. 2, no. 2, pp. 109-114, 2015.
- [16] Park, K., Hong, Y.K., Ki, G.H., and Lee, J., "Classification of apple leaf conditions in hyper-spectral images for diagnosis of Marssonina blotch using mRMR and deep neural network", *Computers and Electronics in Agriculture*, Vol 148, pp 179–187, 2018.
- [17] Pereira, D.R., Papa, J.P., Saraiva, G.F.R., and Souza, G.M., "Automatic classification of plant electrophysiological responses to environmental stimuli using machine learning and interval arithmetic" *Computers and Electronics in Agriculture*, 2018, Vol 145, pp 35 - 42, 2018.
- [18] Rumpf, T., Römer, C., Weis, M., Sökefeld, M., Gerhards, R., and Plümer, L., "Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Galium aparine*", *Computers and Electronics in Agriculture* Vol. 80, pp 89–96, 2012
- [19] Pereira, L. F. S., Barbon Jr., S., Valous, N.A., and Barbin, D. F., "Predicting the ripening of papaya fruit with digital imaging and random forests", *Computers and Electronics in Agriculture*, Vol. 145, pp 76–82, 2018.
- [20] Mohapatra, A., Shanmugasundaram, S., and Malmathanraj, R., "Grading of ripening stages of red banana using dielectric properties changes and image", *Computers and Electronics in Agriculture*, volume 143, pp 100-110, December 2017
- [21] Longsheng Fu, Shipeng Sun, Rui Li, and Shaojin Wang, "Classification of Kiwifruit Grades Based on Fruit Shape Using a Single Camera", *Sensors (Basel)*. July 2016, (online) doi: 10.3390/s16071012.
- [22] Yan, H., Zou, Z., and Wang, H., "Adaptive neuro fuzzy inference system for classification of water quality status", *Journal Environ. Sci.* Volume 22, Issue 12, 2010, pp 1891-1896.
- [23] Guney, K. and Sarikaya, N., "Adaptive Neuro-Fuzzy Inference System for Texture Image Classification", *Int. Conf. Autom. Cogn. Sci. Opt. Micro Electro-Mechanical Syst. Inf. Technol.*, vol. 19, no. 3, pp. 3–7, 2015
- [24] Atashi, A., Nazeri, N., Abbasi, E., and Dorri, S., "Breast Cancer Risk Assessment Using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Subtractive Clustering Algorithm", *Multidisciplinary Cancer Investigation*, vol. 1, no. 2, 2017
- [25] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., pp. 1097–1105, 2012
- [26] Ballard, D.H., *An introduction to natural computation*. MIT press 1999
- [27] McCulloch, W.S. and Pitts, W., 1943. "A logical calculus of the ideas immanent in nervous activity". *The bulletin of mathematical biophysics*, 5(4), pp. 115-133, 1943
- [28] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., "Gradient-based learning applied to document recognition" *Proceedings of the IEEE*, 86(11), pp. 2278-2324 1998
- [29] Ilse, M., Tomczak, J.M. and Welling, M., "Attention-based deep multiple instance learning" *arXiv preprint arXiv:1802.04712*, 2018
- [30] Bengio, Y., Louradour, J., Collobert, R. and Weston, J., 2009, "Curriculum learning" *Proceedings of the 26th annual international conference on machine learning* , pp. 41-48, June 2009
- [31] Hinton, G., Vinyals, O. and Dean, J., "Distilling the knowledge in a neural network". *arXiv preprint arXiv:1503.02531*, 2015
- [32] Bucilua, C., Caruana, R. and Niculescu-Mizil, A., "Model compression", *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 535-541, August 2006
- [33] Malaysia National Cancer Registry report (2012 - 2016), National Cancer Institute, Ministry of Health, 116 pages. <https://codeblue.galencentre.org/2020/01/03/cancer-cases-rise-in-malaysia-chinese-most-prone/#:~:text=%E2%80%9CFor%20the%202012%20to%202016,he%20said%20in%20a%20statement>. accessed on 23 October 2020
- [34] Leong, K.W. *Cancerous Breast Tumour Classification in Mammograms with Assistant-based Reverse Knowledge Distillation*. BSc. Thesis. University of Nottingham Malaysia, 2020
- [35] Bowyer, K., Kopans, D., Kegelmeyer, W.P., Moore, R., Sallam, M., Chang, K. and Woods, K., "The digital database for screening mammography", *Third international workshop on digital mammography*, vol. 58, p. 27, June 1996.
- [36] Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K. and Munishkumaran, S., "Current status of the digital database for screening mammography". *Digital mammography*, pp. 457-460, Springer, Dordrecht 1998.
- [37] Mirzadeh, S.I., Farajtabar, M., Li, A. and Ghasemzadeh, H., "Improved knowledge distillation via teacher assistant: Bridging the gap between student and teacher", *arXiv preprint arXiv:1902.03393*, 2019
- [38] He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition" *CoRR abs/1512.03385*, 2015.

- [39] Retail e-commerce sales worldwide from 2014 to 2023. <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/#statisticContainer>. accessed on 23 October 2020
- [40] Michelle Russell. Global online fashion sales up but sales down amid Covid-19. https://www.just-style.com/news/global-online-fashion-sales-up-but-spend-down-amid-covid-19_id138881.aspx, accessed on 21 May 2021
- [41] Bye, E., & LaBat, K. "An analysis of apparel industry fit sessions. *Journal of Textile and Apparel*", Technology and Management, 4(3), pp 1-5, 2005
- [42] DesMarteau, K. Pre-production & CAD: Let the fit revolution begin. *Bobbin*, 42, pp 42–56, 2000.
- [43] Goldsberry, E., Shim, S., & Reich, N. "Women 55 years and older: Part II. Overall satisfaction and dissatisfaction with the fit of ready-to-wear", *Clothing and Textiles Research Journal*, 14(2), pp 121-132, 1996
- [44] Li, M., and Li, H. "Application of deep neural network and deep reinforcement learning in wireless communication", *PloS one*, 15(7), 2020, e0235447. <https://doi.org/10.1371/journal.pone.0235447>.
- [45] Stankov, S., Ivanov, S., and Todorov, T. "An application of deep neural networks in industrial robotics for detection of humans", *IEEE XXVIII International Scientific Conference Electronics (ET)*, Sozopol, Bulgaria, pp. 1-3, doi: 10.1109/ET.2019.8878583, 2019
- [46] Wang, J., Ma, Y., Zhang, L., Gao, R., and Wu, D. "Deep learning for smart manufacturing: methods and applications", *Journal of Manufacturing Systems*. 48. pp 144-156., 2018 10.1016/j.jmsy.2018.01.003.
- [47] Villalba-Diez, J., Schmidt, D., Gevers, R., Ordieres-Meré, J., Buchwitz, M., and Wellbrock, W. "Deep learning for industrial computer vision quality control in the printing industry 4.0". *Sensors (Basel, Switzerland)*, 19(18), 3987, 2019. <https://doi.org/10.3390/s19183987>.
- [48] Julia M. H., Noothout, Bob D. de Vos, Jelmer M. Wolterink, Elbrich M. Postma, Paul A. M. Smeets, Richard A. P. Takx, Tim Leiner, Max A. Viergever and Ivana Išgum. "Deep learning-based regression and classification for automatic landmark localization in medical images". *arXiv*. <https://arxiv.org/pdf/2007.05295.pdf>, 2020
- [49] Honari, S., Molchanov, P., Tyree, S., Vincent, P., Pal, C., & Kautz, J. "Improving landmark localization with semi-supervised learning", 2018. *arXiv*. <https://arxiv.org/abs/1709.01591>.
- [50] Feng, Z., Kittler, J., Awais, M., Wu, X. "Rectified wing loss for efficient and robust facial landmark localisation with convolutional neural networks". *Int J Comput Vis* 128, 2126–2145, 2020. <https://doi.org/10.1007/s11263-019-01275-0>.
- [51] Dong, X., and Yang, Y. "Teacher supervises students how to learn from partially labeled images for facial landmark detection", *The International Conference on Computer Vision*. https://openaccess.thecvf.com/content_ICCV_2019/papers/Dong_Teacher_Supervises_Students_How_to_Learn_From_Partially_Labeled_Images_ICCV_2019_paper.pdf, 2019
- [52] Song, Y., Qiao, X., Iwamoto, Y., and Chen, Y.-W. "Automatic cephalometric landmark detection on x-ray images using a deep-learning method" *Applied Sciences*. 10.2547.10.3390/app10072547 , 2020
- [53] Chen, Y., Tong, Z., Zheng, Y., Samuelson, H., and Norford, L., "Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings" *Journal of Cleaner Production*, 254, 2020 <https://doi.org/10.1016/j.jclepro.2019.119866>
- [54] Shorten, C., Khoshgoftaar, T.M. "A survey on image data augmentation for deep learning" *J Big Data* 6, 60, <https://doi.org/10.1186/s40537-019-0197-0> 2019
- [55] Liu, Z., Luo, P., Qiu, S., Wang, X., and Tang, X., "Deepfashion: powering robust clothes recognition and retrieval with rich annotations" *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016.
- [56] Mykhaylo, A., Leonid, P., Peter, G., and Schiele, B. "2D human pose estimation: new benchmark and state of the art analysis" *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [57] Charles, J., Pfister, T., Magee, D., Hogg, D., and Zisserman, A. "Personalizing human video pose estimation", *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [58] Liu, J., and Lu, H., "Deep fashion analysis with feature map upsampling and landmark-driven attention" *Proceedings of European Conference on Computer Vision*. 30-36, 2018
- [59] Hermawan, E. S., Human body landmark detection for customised fashion design. Final Year Project Dissertation, University of Nottingham Malaysia, 2020.
- [60] He, K., Zhang, X., Ren, S., and Sun, J. "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification" *IEEE International Conference on Computer Vision (ICCV)*, Santiago, pp. 1026-1034, 2015 doi: 10.1109/ICCV.2015.123.
- [61] Kingma, D., and Ba, J. "Adam: a method for stochastic optimization" *International Conference on Learning Representations*, 2014 *arXiv*: <https://arxiv.org/pdf/1412.6980.pdf>.

PROFILES



KAM MENG GOH received both the bachelor degree and the Ph.D. degree from University Teknologi Malaysia in 2010 and 2015 respectively. He is currently a senior lecturer as well as programme leader with Tunku Abdul Rahman University College in Kuala Lumpur, Malaysia. His current research interests include computer vision, machine vision and artificial intelligence. He has published more than 15 international peer reviewed papers. He is also recently nominated as a senior member in IEEE, and Associate Fellow of Asean Academy of Engineering and Technology (AAET).
Email address: gohkm@tarc.edu.my



ERIC SAVERO HERMAWAN received a BSc (Hons.) and S. Kom from University of Nottingham, Malaysia and Binus International, Indonesia respectively in 2020. Also in 2020, he did his internship as a lab assistant in a machine learning project. Currently he is working as a data analyst in Cakap, an Indonesian E-learning company.
Email address: eric.savero9092@gmail.com



MEI YUAN KOAY received the B.Eng Degree (Hons) in Electrical & Electronics Engineering from Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia in 2017. Mei Yuan is a recipient of the IEM Gold Medal Award in 2017, special prize for IEM Best Final Year Project in 2018 and Best Presenter at the 7th International Conference on Control, Robotics and Informatics (ICCRI). Currently she is an engineer with Intel (Malaysia). Her research interests are in computer vision and artificial intelligence.
Email address: meiyuan72@gmail.com



WENG KIN LAI is an Associate Professor with the Department of Electrical & Electronic Engineering, Faculty of Engineering & Technology, Tunku Abdul Rahman University College. He received his Doctorate (by research) in Engineering from the University of Auckland, New Zealand. Weng Kin is a Fellow of the IET, registered Chartered Engineer with the Engineering Council of U.K., senior member of IEEE, Associate Fellow of AAET, and a Professional Engineer, registered on the International Professional Engineer register (U.K.). He is a Governing Board member of the *Asia Pacific Neural Network Society* since 2014. His research interests are in computer vision, artificial intelligence and their applications.
Email address: laiwk@tarc.edu.my



IMAN YI LIAO obtained her BSc, MSc, PhD all from School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, China, in 1996, 1999, and 2006 respectively. She was a visiting researcher in Chinese University of Hong Kong from July 2001 to Dec 2001, and in CVSSP, University of Surrey, UK from Dec 2003 to Sept 2004, respectively. She joined the School of Computer Sciences, Universiti Sains Malaysia in 2008 as a Post-Doctorate research fellow and subsequently became a Senior Lecturer there until 2012. She is now an Associate Professor in School of Computer Science, University of Nottingham Malaysia Campus.
Email address: Iman.Liao@nottingham.edu.my



KAR WEY LEONG obtained a BSc (Hons.) degree in Computer Science from the University of Nottingham Malaysia, and is currently a Product Manager at Setel. Kar Wey has experience in the research and development of deep learning models, particularly in the area of cancer diagnosis.
Email address: karwey0807@gmail.com



XIU KAI LIM received the B.Eng Degree (Hons) in Electrical & Electronics Engineering from Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia in 2018. Currently he is an engineer with Gin Seiko Precision Sdn. Bhd. His research interests are in computer vision and artificial intelligence.
Email address: melvin55.ml@gmail.com



SELINA LOH received the B.Eng Degree (Hons) in Mechatronics Engineering from Tunku Abdul Rahman University College, Kuala Lumpur, Malaysia in 2017. Currently she is an engineer with Mi Equipment. Her research interests are in computer vision and artificial intelligence.
Email address: selinaloh_810@hotmail.com



TOMÁS MAUL received a BSc (Hons.) degree in Psychology from the University of St. Andrews, St. Andrews, UK, two MSc degrees in Computer Science from Imperial College, London, UK, and a PhD degree in Computer Science (Computational Neuroscience) from the University of Malaya, Kuala Lumpur, Malaysia. For two years, he was a Senior Researcher with MIMOS Berhad, where he worked in the fields of pattern recognition and computer vision. He is currently Associate Professor at the School of Computer Science, University of Nottingham Malaysia, where he lectures and conducts research mainly in the area of neural computation
Email address: Tomas.Maul@nottingham.edu.my