OPTIMISING IMAGE CLASSIFICATION

Implementation of Convolutional Neural Network Algorithms to Distinguish Between Plans and Sections within the Architectural, Engineering and Construction (AEC) Industry

JENNIFER MEI YEE NG[1], NARIDDH KHEAN[2],
DAVID MADDEN[3], ALESSANDRA FABBRI[4],
NICOLE GARDNER[5], M. HANK HAEUSLER[6] and
YANNIS ZAVOLEAS[7]
[1,4,5,7] *UNSW*
[1,4,5,7] *{j.ng|a.fabbri|n.gardner|y.zavoleas}@unsw.edu.au*
[2,3] *Arup Engineering Sydney*
[2,3] *{Nariddh.Khean|David.Madden}@arup.com*
[6] *UNSW & CAFA Visual Innovation Institute Beijing*
[6] *m.haeusler@unsw.edu.au*

Abstract. Modern communication between built environment professionals are governed by the effective exchange of digital models, blueprints and technical drawings. However, the increasing quantity of such digital files, in conjunction with inconsistent filing systems, increases the potential for human-error upon their look-up and retrieval. Further, current methods are manual, thus slow and resource intensive. Evidently, the architectural, engineering and construction (AEC) industry lacks an automated classification system capable of systematically identifying and categorising different drawings. To intercede, we aim to investigate artificially intelligent solutions capable of automatically identifying and retrieving a wide set of AEC files from a company's resource library. We present a convolutional neural network (CNN) model capable of processing large sets of technical drawings - such as sections, plans and elevations - and recognise their individual patterns and features, ultimately minimising laboriousness.

Keywords. Convolutional Neural Network; Artificial Intelligence; Machine Learning; Classification; Filing architectural drawings.

1. Introduction

Through popular accounts, artificial intelligence (AI) is often portrayed as the source of futuristic dystopic outcomes. However, machine learning (ML), a sub-field of AI, has been progressively integrated into our daily life with positive outcomes. From the micro-scale of an individual's smartphone, home and security systems and personalised social media feeds, to the macro-scale of urban transportation systems and cyber-physical technologies, ML has contributed to

Intelligent & Informed, Proceedings of the 24th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2019, Volume 2, 795-804.

lessening repetitive manual tasks while improving the quality of life. Within the built environment, ML has shown potential in implementing traditional design processes and approaches. By way of explanation, in 2016, Phelan investigated employees' dynamics in a workplace with the goal of developing a more efficient work environment. After collating the usage data of the workspaces, a ML algorithm was implemented to predict rooms occupancy, ultimately to offer the most appropriate space for an improved working experience. Correspondingly, the present research aims to investigate workplace efficiency with the aid of ML. As architectural drawings have entered the digital realm, a variety of files progressively accumulates in local and global servers. These files are often unstructured and unsystematic, forcing engineers and designers to sort and retrieve them manually and repetitively. With ML coming into place, engineers can be off-loaded of laborious duties, while intelligent alternatives can be developed.

2. Research Aims

Challenging traditional approaches to architecture, engineering and design, the present research aims to investigate the potential of ML in optimising workplace efficiency by constructing a predictive model that can differentiate between typologies of architectural drawings (i.e. plans and sections). The initial objective is to master our understanding of ML concepts and principles, while determining what algorithm is best suited to solve the problem at hand. The second objective is to construct a model that effectively recognises and categorises given images, and to do so under optimal hyperparameters.

3. Research Question

Ultimately, the research project revolves around the following question: *"How can drawing classification be automated within the built environment discipline using machine learning?"*

Subsequently, several sub-questions emerge from the iterative nature of the investigative process, such as *"How effective is the convolutional neural network algorithm for the given problem?"*; and *"What are the differences between 'training', 'validation' and 'prediction', and how are they indicative of a ML model's performance?"*; and *"How can the adjustment of CNN hyperparameters - user controlled variables - affect the model's accuracy?"*.

4. Methodology

An action research approach is held to carry out the research. Action research, as per D. Gabel's definition (1995), involves a heuristic cycle, which iteratively progresses through the conceptualisation of a problem (exploratory stance), the action towards its resolution (action), and an evaluation of that action (observation) (Khean et al., 2018). In the first stage (exploratory stance), the impact of AI and ML on the architectural, engineering and construction (AEC) industry are investigated to highlight the need of ML for an intelligent and informed approached to problem-solving. In the second stage (action), ML fundamentals are explored to identify the best algorithmic approach for the current investigation. Subsequently, a predictive model is constructed. In the third stage (observation), the predictive model is tested, quantitative performance feedbacks are collated, and observations are produced in concert with an industry partner (Arup Engineering), academic supervisors, and colleagues. Lastly, the process is iteratively repeated until the desired outcome is achieved. Observations and feedbacks are directed to always reflect the following questions: Can the model be adapted to different drawing style? Can the model differentiate between line thicknesses, colours and texts? Finally, can the model be implemented to accommodate different applications inside and outside the discipline of built environment?

5. Background Research

EVOLUTIONAL CHANGE IN ARCHITECTURAL DESIGN PROCESSES Since the historical shift between the Renaissance master builder and the modern architectural profession, design processes have continued to evolve with the fast progression of technological advancements. Novel computational tools have emerged that have allowed for exploring uncharted territories, where manual drawings have been progressively replaced by digitally-aided methods. According to Wirz's *"revolutionary cycle"* (2016), this metamorphosis has yet to reach full completion. However, an interesting trend has already risen. Designers and researchers have not only rapidly engaged in the exploration of advanced solutions for complex designs, but they have also increasingly shared their experience and knowledge through open-source platforms. With novel computational methods and approaches being more and more accessible to the wider community, students have been provided with the opportunity to rapidly understand this fundamental change in the design paradigm, where complexity and performance are explored not as separate aspects from their application in a specific form, but rather as complex interrelations between materials, structure and behaviour. In 1987, the architect Peter Eisenman was appointed by the Goethe University of Frankfurt for the design and construction of the Biozentrum, namely a biological centre which included biotechnology, molecular biology, and biochemistry research laboratories and support spaces. Eisenman's design inspiration came from the complexity of biological processes such as the human DNA, which he wished to embed, embody and celebrate within a long-lasting monument. The apparently aesthetic direction of Eisenman's design concept shouldn't, however, deceive the reader. Biological theories and procedures were mainly employed to optimise the distribution, dimension and form of the centre's innumerable spaces. To address such complexity of interrelations, Eisenman turned to computing, and initiated the development of ”a procedural modelling tool capable of drafting predefined figures at varying alignments” (Zardini, 2013). The resulting geometrical pattern reflected a fractal distribution of individual laboratories, where varying scales and oblique orientations served to generate cantilevered spaces for cafeteria, library and meeting functions. As we briefly delineated with Eisenman's example, the increasing variation in computational methods has proved to be invaluable not only to achieve purely aesthetic outputs but also to optimise functionality and technicality of complex designs. A fruitful cointegration between aestheticism and functionalism requires, however, a certain cross-disciplinary approach, where disciplines such as engineering, architecture, biology, and computer science cohesively contribute to the design process. With consideration for the aforementioned, the current research aims to refine and enhance a specific design process (i.e. files' sorting and identification) through the exploration and validation of a novel cross-disciplinary approach. Ultimately, we hope to encourage experimentation outside the box and further optimisation of design solution where form is informed by performance.

ML AND THE CHANGING PARADIGM Entering the lexicon in 1959, the term "machine learning" has continuously evolved, still failing to reach an agreed upon definition today. However, Professor Tom M. Mitchell of Carnegie Mellon University eloquently describes that "a machine learns with respect to a particular task, performance metric, and type of experience, if the system reliably improves its performance at that task, following experience" (Broussard, 2018). Within this paradigm, artificial neural networks (ANNs), a subcategory of ML algorithms, have been successfully applied to recognising patterns and trends within a specific dataset. Originally inspired by the human neurological system, ANNs can mimic our brain's information processing and knowledge acquisition - initially starting at a randomised state, and improving its performance over time through experience. In 2000, Basheer and Hajmeer published the article *Artificial Neural Networks: Fundamentals, Computing, Design, and Application*, where the juxtaposition between ANN approaches and traditionally held techniques was analysed. While highlighting the higher degree of robustness and rapidity of ANNs, the authors also introduced the idea of automation of design practices with the goal of expanding design opportunities, within and outside the design realm. A variation of the ANN algorithm is the convolutional neural network (CNN). CNNs differ from the ANN architecture due to its inherent ability to comprehend spatial relationships. There have been several instances wherein CNNs have been implemented into processes involving image recognition, semantic segmentation, and even image manipulation and generation. In 2015, Gatys, Ecker and Bethge investigated the potential of style transferring between several pictures. As such, they first dismantled the images in hierarchical layers. Subsequently, they identified specific features in each image that could reveal and transpose the styles from one another. With the overarching objective of developing a computational framework for recognising and categorising architectural drawings, we believe that an ANN, and in particular a CNN, would suit the present research. However, before constructing a model, we thought necessary to obtain a comprehensive understanding of how CNNs behave and what is their fundamental structure. In 1980, Fukushima published a paper titled Neocognitron: A Self-Organising Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shifts in Position. The article focused on the improvement of a model's ability to recognize patterns, namely neocognitron, which required an extended self-organizing multi-layered neural network. The structure proposed by Fukushima resembled the hierarchical model of the visual nervous system developed by Hubel and Wiesel between 1962 and 1965, where information flowed from simple cells to complex cells. In Fukushima's proposal, the to-be-improved network was aimed to implement position variations within input data. Previous research had demonstrated the sensitivity of such data to normalisation. Fukushima implemented a position-invariant pattern recognition, where the network's structure consisted of C and S cell layers sub-grouped into ”cell-planes”. Each cell would have fields receptive of the same function independent of their position, and with the layers receiving the same set of input synapses, parameter alternations would provide the model with the determinants for differentiating between patterns similarities and distortion. Fukushima's network progressively acquired the ability to correctly distinguish between different inputs with similar features. However, when prompted to increase the number of cell-planes with the goal of increasing the steadiness and accuracy of the model's prediction, the network failed to perform due to a computer's memory deficiency. To achieve the overarching aim of the present research, the authors believe in employing Fukushima's network and method for position-invariant recognition. Since Fukushima's research, computers' memory has increased exponentially, allowing us to progress from where he left.

6. Case Study

In collaboration with Arup Engineering, an independent firm of multidisciplinary consultants in the field of Built Environment, this research aims to develop, train and validate a neural network model that can recognise and categorise different drawings systematically. To do so, an application-based approach is pursued. A CNN model is implemented with the goal of identifying the spatial relationship between pixels of the same image, the structure and behaviour of which we delineate below.

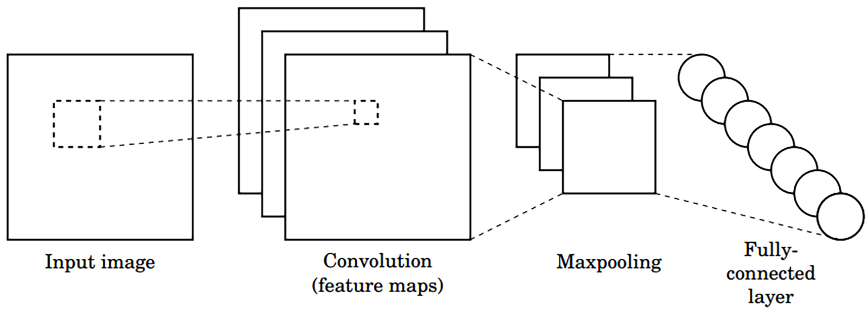

Figure 1. Diagram of Convolutional Neural Network (Pavlovsky, 2017).

INPUT LAYER In the first stage, images of the same category (either plans or sections) are fed to the model for training and validation. Each image is broken down by the algorithm into a matrix of input values, where each value corresponds to a pixel or, more specifically, to a combination of the pixel's RGB codes. In the process, 800 images are inputted for training and 200 images for validation.

CONVOLUTIONAL LAYERS In the second stage, each pixel is analysed by a set of filters, known as *kernels*. Each kernel, initialised randomly, moves along the input matrix of pixels by *n* strife steps. At each step, the kernel overlaps a portion of the image's pixels, which are multiplied and summed, returning a single value. As the kernel moves along the input image, a single value is returned, leading to a new matrix for the next layer.

MAXPOOLING Following the generation of the new matrix, the image size is reduced in the third stage. Observing a portion of the matrix at a time, the largest value in the portion is kept and relatively insignificant details are ignored, subsequently reducing training time and preventing overfitting.

FULLY CONNECTED LAYERS In the last stage of the model, the matrix is flattened, while every value previously calculated is connected with the scope of training, validating and predicting the output data. The aim of the previous layers were to extract the features of the image, while the fully-connected layer is designed to map the correlation between these extracted features and the image's classification.

TRAINING AND VALIDATION The results from the training and validation process, indicated by the model's accuracy, indicates how best to modify the CNN's hyperparameters. The goal of this iterative adjustment is to increase the model's ability to generalise the correlation between features and image classes, which in turn increases the model's robustness. As a result, many iterations are generated, of which three are selected and evaluated in detail in this paper. Hence in the following we are changing the hyperparameters.

```
Number of Convolutional Layers: 1
Number of Filters: 32
Filter Size: (3, 3)
Dropout Percentage: 0.2
Learning Rate: 0.001
Epooch: 128
Batch Size: 8
```

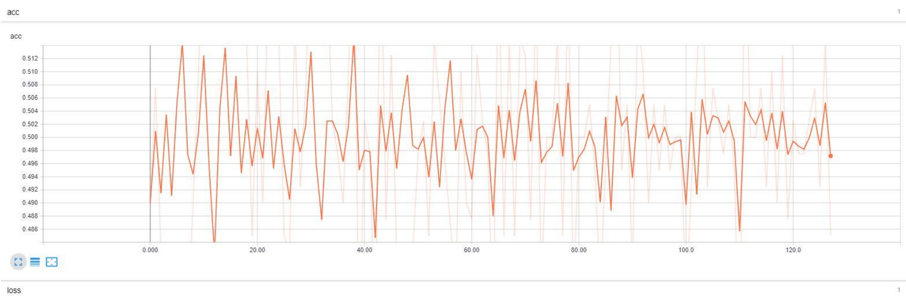

Figure 2. Iteration 1, Training Accuracy.

The performance of the model can be evaluated through the model's accuracy, calculated by comparing the number of correctly classified images against those classified incorrectly. For this research, the objective was to achieve an accuracy of over 90%. In theory, as the model continues to train, its accuracy should improve. However, poorly chosen hyperparameters will manifest as excessive fluctuations or even a decrease in accuracy, also known as “overfitting”. Overfitting is an undesirable consequence of a model with ill-fitting hyperparameters and can be intuitively described as the memorisation of the training data instead of the generalisation of the underlying patterns and trends. There are loose guidelines for selecting optimal hyperparameters, however each problem requires some level of hyperparameter tuning. Below are the hyperparameters of Iteration 1:

Analysing the Iteration 1 training graphs (Figure 2), we can observe that the model’s accuracy reached a peak accuracy of 52%. Considering the task is a binary classification task, where a purely random classifier would typically get 50% accuracy, a 2% improvement does not indicate effective learning. However, it is the sporadic nature of the graphs are more telling. Changes this severe aren’t typical of effective training. A likely reason for this is because the model is not recognising enough features from the input images. To improve the learning process, two additional layers are added to the convolutional network, and Iteration 2 is initiated. Below are the hyperparameters of Iteration 2:

```
Number of Convolutional Layers: 3
Number of Filters: 32
Filter Size: (3, 3)
Dropout Percentage: 0.2
Learning Rate: 0.001
Epooch: 128
Batch Size: 8
```

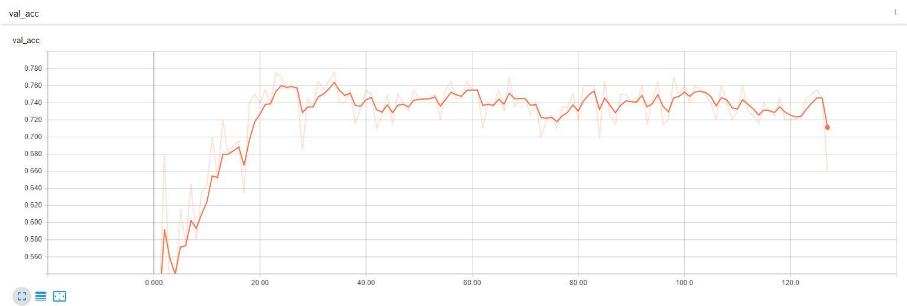

Figure 3. Iteration 2, Validation Accuracy.

Increasing the number of convolutional layers, which increases the complexity of extracted features, greatly reduced the extreme changes in accuracy. This smoother change is indicative of gradual improvement, thus more effective learning. Iteration 2 is a step in the right direction, however still shows larger fluctuations near the end of the 128 epochs - potential signs of overfitting. Accordingly, the model’s learning rate is lowered in the following iteration. Below are the hyperparameters of Iteration 3:

```
Number of Convolutional Layers: 3
Number of Filters: 32
Filter Size: (3, 3)
```

```
Dropout Percentage: 0.2
Learning Rate: 0.0001
Epooch: 128
Batch Size: 8
```

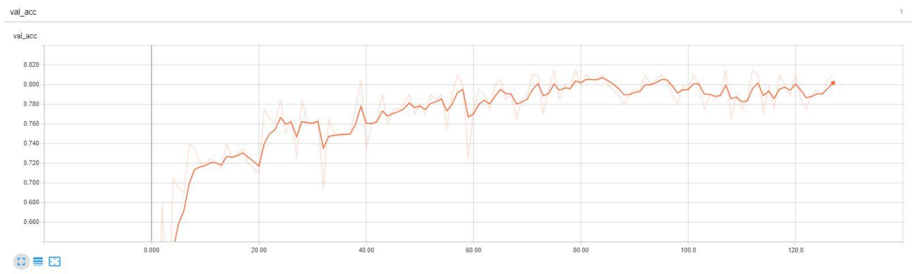

Figure 4. Iteration 3, Validation Accuracy.

In Iteration 3, the model achieves a semblance of stability, indicated by more modest increases in accuracy. Despite our initial aim for a model with greater than 90% accuracy, the state of the model at its peak accuracy of 82% would suffice. It is this model that is held for prediction.

PREDICTION Following a dual-sample testing approach, the model's robustness is tested against new images that the model was not trained with. Because of the unicity of the architectural form, six Sydney Opera House plans and sections were used to assess if the trained model has truly learnt to extract the characteristics of what makes a plan and section. The model successfully classified all of these images (Figure 5), however, failed to achieve this same accuracy when tested on further inputs. As such, we recommend further adjustment of the model's hyperparameters before its implementation within a practice-based scenario.

7. Significance of Research

The present research aimed to investigate the development of a artificially intelligent system, capable of identifying and categorising a wide set of AEC files. The project was developed to encourage greater efficiency in the workplace by offloading employees of time-consuming and laborious tasks, such as the manual sorting of architectural drawings. A CNN model was developed and trained to process an abundance of imagery files and successfully classifyed them into their respective categories with an accuracy of 82%. The immediate next steps would involve further iterations of hyperparameter adjustment (with longer training times), in an attempt to achieve the desired 90% accuracy. This would be greatly aided with a larger training dataset of pre-classified drawings, as well as explorations of new CNN architectures - such as highway networks (Srivastava et al. 2015), residual networks (He et al. 2016), and dense networks (Huang et al. 2017) - that have been shown to yield superior performance. Assuming a trained model with acceptable levels of accuracy, the final steps required before deploying the tool would involve the development of an interface accessible to non-experts. This could take the form of a program or an executable file (py2exe), which professionals can run locally. Alternatively, the trained model can be added to a backend server and uploaded to the cloud as a web application.

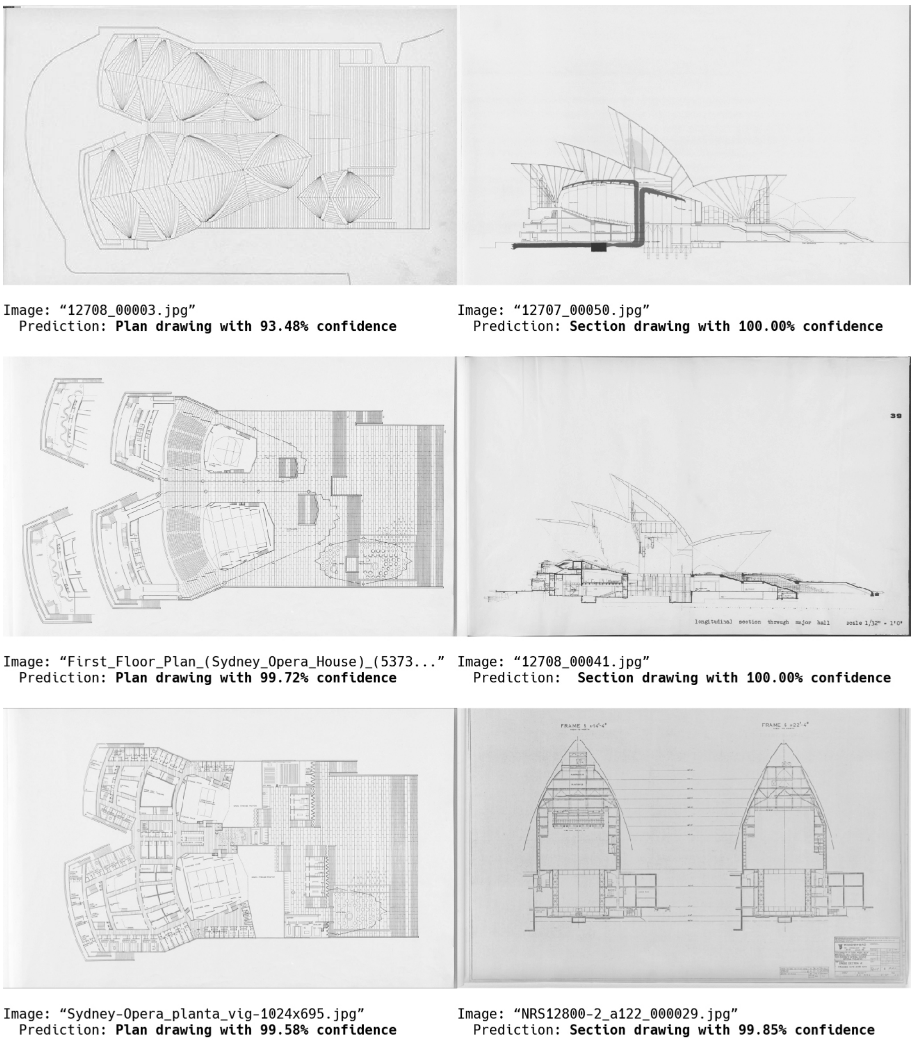

Figure 5. Original Sydney Opera House Drawings and Predictions.

8. Conclusions

Workplace efficiency can and should be addressed with the use of algorithmic solutions such as the one proposed in the present paper. Clearly, we aim to encourage a better communication between professionals of the AEC industry through an enriched organisation of resources libraries, and to increase their productivity. The model presented in this paper could be further implemented to orderly categorise digital or hand-drawn archival files which might be currently forgotten in the companies' storage, and which could still inspire knowledge and innovation. Moreover, the model shows potential for identifying more specific features such as abstract drawings, texts and renders, to eventually differentiate between real pictures and architectural visualisations of the same building.

References

"A simple way to understand machine learning vs deep learning" : 2018. Available from Zendesk<https://www.zendesk.com/blog/machine-learning-and-deep-learning/> (accessed 22 September 2018).

Basheer, I. and Hajmeer, M.: 2000, Artificial neural networks: fundamentals, computing, design, and application, *Journal of Microbiological Methods*, **43**(1), 3-31.

Broussard, M.: 2018, *Artificial Unintelligence: How Computers Misunderstand the World*, The MIT Press, United States of America.

Fukushima, K.: 1980, Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position, *Biological Cybernetics*, **36**, 193 - 202.

Gatys, L. A., Ecker, A. S. and Bethge, M.: 2015, A Neural Algorithm of Artistic Style, *CoRR*, **1**, 1 - 16.

He, K., Zhang, X., Ren, S. and Sun, J.: 2016, Deep Residual Learning for Image Recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 770-778.

Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K.Q.: 2017, Densely Connected Convolutional Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 4700-4708.

Khean, N., Kim, L., Martinez, J., Doherty, B., Fabbri, A., Gardner, N. and Haeusler, M. H.: 2018, The Introspection of Deep Neural Networks - Towards Illuminating the Black Box, *Proceedings of CAADRIA 2018*, Beijing, 237 - 246.

Pavlovsky, V.: 2017, "Introduction To Convolutional Neural Networks" . Available from <https://www.vaetas.cz/posts/intro-convolutional-neural-networks/> (accessed 9th December 2018).

Phelan, N.: 2016, "Designing with Machine Learning" . Available from <https://www.wework.com/blog/posts/designing-with-machine-learning> (accessed 1st November 2018).

Shah, D.: 2018, "AI, Machine Learning, & Deep Learning Explained in 5 Minutes." . Available from Becoming Human: Artificial Intelligence Magazine.<https://becominghuman.ai/ai-machine-learning-deep-learning-explained-in-5-minutes-b88b6ee65846> (accessed 20 September 2018).

Sharma, A.: 2017, "Convolutional Neural Networks in Python." . Available from DataCamp Community<https://www.datacamp.com/community/tutorials/convolutional-neural-networks-python> (accessed 22 September 2018).

Srivastava, R.K., Greff, K. and Schmidhuber, J.: 2015, Training Very Deep Networks, *Proceedings of Advances in Neural Information Processing Systems*, Montréal, 2377-2385.

Wirz, F. 2016, Introduction, *in* A. Tedeschi (ed.), *AAD Algorithms-Aided Design. Parametric strategies using Grasshopper.*, Le Penseur, Brienza, 10 - 11.

Zulkifli, H.: 2018, "Understanding Learning Rates and How It Improves Performance in Deep Learning." . Available from <https://towardsdatascience.com/understanding-learning-rates-and-how-it-improves-performance-in-deep-learning-d0d4059c1c10> (accessed 9th December 2018).