A Model Approach for an Automatic Clothing Combination System for Blind People

Daniel Rocha[1(✉)], Vítor Carvalho[1,2], Filomena Soares[1], and Eva Oliveira[2]

[1] Algoritmi R&D, University of Minho, Guimarães, Portugal
id8057@alunos.uminho.pt, fsoares@dei.uminho.pt
[2] 2Ai Lab, School of Technology, IPCA, Barcelos, Portugal
{vcarvalho,eoliveira}@ipca.pt

Abstract. To dress adequately may be a necessary condition in social interaction. The way we dress may have an impact in the way people see us. Recognizing and matching clothes in order to dress properly can be a hard and daily stressful task for blind people. How do they recognize and identify the garments attributes to perform an outfit without help? In order to overcome this stressful situation, we present a project to help blind people in the identification and selection of garments.

Keywords: Convolutional neural networks · Blind people · Clothing recognition

1 Introduction

More and more, aesthetics is present in our lives. The way we dress may have a large impact on our daily lives and may be critical to our well-being.

In fact, our appearance has an impact in the way people see us and the perception they create of us can have a huge influence on the workplace. Some particular events imply dress code and special dedication is devoted to select and combine garments, taking care of all the details from the colours, to the textures, fabrics, accessories, shoes, and scarves. The way we dress is sometimes considered as a business card. This task may be particularly difficult for visually impaired persons, especially if they do it by themselves.

Blind people are dependent on relatives (family, friends) to buy clothes, to detect dirt or to choose the colour of the clothes they want to wear. These are not difficult tasks for those who see, but for those who have little or no vision they become extremely difficult.

For blind people recognizing and choosing an outfit, in order to dress properly, can be a difficult and daily stressful task. Following the previous work [1, 2], this paper is focused in the attempt to recognizing the attributes of the garments to perform later an automatic combination clothing system for blind people.

As mentioned in [2], extracting the clothes features is essential to achieve the goal of automatic combinations. In this sense, in this paper we present the starting point of extracting and classifying the clothing type from an image.

This paper in divided in five sections. Section 2 describes the related work; in Sect. 3 we present the project overview; Sect. 4 explains the methodology, Sect. 5 describes the experiments and finally, Sect. 6 concludes with the final remarks.

2   Related Work

Some fashion apps have been developed as STYLEBOOK, an application to manage the clothes, create outfits, and plan what to wear [3]. ShopStyle [4] allows to plan purchases based on the user's favourite stores, searching items across the web. Another application is Tailor [5] that is a closet that learns the user's preferences and choices and suggests the combinations to wear.

Electronic devices such as Colorino has come to fill the difficulties of the blind in the distinction of colours for the most varied tasks, since it helps in the choice of clothes3, the washing procedure and the colour combination [6]. Another device is the ColorTest 2000, which is a device similar to Colorino that identifies colours and reads the date and time and detects if a light is switched on or off [7].

Yamaguchi et al. [8, 9] propose an approach to clothe analysis. The analysis approach consists in retrieving similar images from a predefined database. The proposed clothing analysis solution is able to classify 53 different categories of fashion photo clothes. The method is able to separate segments in each piece of clothing.

Wazarkar and Keshavamurthy in [10] propose the classification of fashion images, incorporating the concepts of linear convolution and corresponding points using local characteristics. Another study proposes a system for the recognition of patterns in clothes and the dominant colour in each image. The system is a finger-based camera that allows users to query clothing colours and patterns by touch [14].

Yang and Yu [11] propose a real-time clothing recognition method in surveillance settings recognizing eight clothing categories using Histogram of Gradients (HOG) with linear Support Vector Machine (SVM) classifiers.

Yuan, Tian, and Arditi [12] developed a prototype based on computer vision to combine a pair of images of two clothes for pattern and colour. The proposed method for pattern detection achieved 85% accuracy, being robust for clothes with complex texture patterns, various colours and variations in rotation and lighting. For pattern matching, the main errors occurred with images with very similar texture patterns. Regarding colour matching in 100 pairs of images, only one pair did not match correctly due to background distraction. In order to deal with complex texture patterns and lighting changes, they combined techniques using Radon transform, wavelet features, and co-occurrence matrix for pattern matching. The results of the assessment of clothing datasets demonstrate that the method is robust and accurate for clothing with complex patterns and various colours. The corresponding outputs are provided to the user in audio.

Yang, Yuan, and Tian [13] proposes a system for recognizing clothing patterns. This system can identify 11 colours of clothing and recognize 4 categories of clothing patterns. A prototype was developed, based on a camera incorporated in the glasses, a microphone, a computer, and a Bluetooth headset for describing clothing patterns and colours.

The deep learning models dedicated to fashion models prove the importance of neuronal networks in this area. As examples of these works, it is important to refer [14, 15] where the fashion network is based on the VGG-16 architecture. Simonyan and Zisserman [16] introduced a VGG network where they evaluated very deep convolutional neuronal networks for large scale image classification concluding that the depth is beneficial for classification accuracy.

In [17] it is presented a prototype system of clothes detection and classification based on convolutional neural networks. They exceeding F-Score of 0.9 in clothes detection accuracy for five major clothes types.

The work in [18] proposes a method to detect fashion items in a given image using deep convolutional neural networks, achieving 86,7% the average in recall.

Although there are already some studies dedicated to fashion models, there is not yet a solution focused in the development of an automatic system for combining and identifying the garments for blind people.

Considering the research previously carried out, there are tools to directly or indirectly identify colours, patterns, garments but there is not yet available an automatic system for combining and identifying clothing items for the blind. There is where this project stands. The objective is to develop a physical prototype capable of storing garments that identifies the clothes and their wear, dirt, colours and patterns, having the ability to independently suggest clothing combinations to the user.

3   Project Overview

This project follows the work previously carried out, where a combination clothing system for the blind people was created, based on NFC technology (Near Field Communication), placing an identification tag on the garment. With the help of a web application, the blind was able to identify the characteristics of the clothes by reading the label, in addition to managing his/her clothes and combinations[19–21].

The aim now is to introduce artificial intelligence algorithms to suggest combinations and extract characteristics from garments. In order to address this issue of the combination of clothing for the blind people, the development of an autonomous system with artificial intelligence is the basis for the solution to this problem.

As there is still no support system for the blind people, developing a mechatronic prototype system with artificial intelligence will help providing independence and consequent well-being in the identification and combination of clothes, contributing to fill the gap of a technological lack in terms of aesthetics and image of a blind person.

The objective of this work is to develop a prototype that recognizes and takes into account the following elementary requisites:

- Type of clothes;
- The season of the year and weather;
- Suggesting combinations of clothing pieces;
- Identification of the clothe pattern;
- Presence of stains;
- Modifications in the state of the garment.

In pursuit of the objectives, the research question of the overall project arises:

- How can a mechatronic device with artificial intelligence make the inspection, identification, combination and management of clothing for a blind person?

It is worth mention that this work has the collaboration of the Association of the Blind and Amblyopes of Portugal (ACAPO). The main objective of the partnership is to design, enhance and validate all the work that is being developed.

4   Methodology

The literature review allowed to identify the research opportunity, as the existing technologies and investigations are few and limited.

In order to achieve the proposed objectives, a deductive approach is made, formulating hypotheses based on a critical analysis of the literature review, since it is based on scientific principles.

The research strategy adopted, the action research is in accordance with the research question and the final objective, in which a prototype must be reached, Fig. 1.
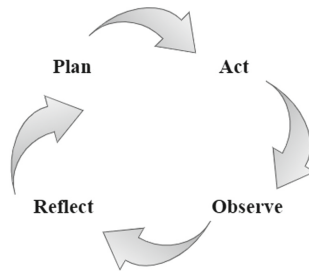

Fig. 1. Action research cycle.

Given that it is intended to develop a technology with practical application in the real context, this strategy is supported by active research, with the participation and collaboration of the target audience.

The process of completing the prototype includes at least two iterations, meaning that two cycles from the beginning to the end are needed to obtain significant improvements. It means each time a cycle is repeated the set of improvements introduced, in order to obtaining the objectives initially stated, tends to stabilize. Qualitative and quantitative data are used to assess the usability and performance of the prototype.

In the development of the tool for image acquisition at an early stage, a mobile device will be used due to accessibility and usability.

The output from an artificial intelligence algorithm will feed the entire system in order to create a database. To obtain the combinations we need to characterize and recognize the characteristics of the garments, as well as, to have a description of the garments contained in the wardrobe. The latter is considered important for the user.

The entire process from acquiring an image to obtaining a combination is illustrated in Fig. 2.
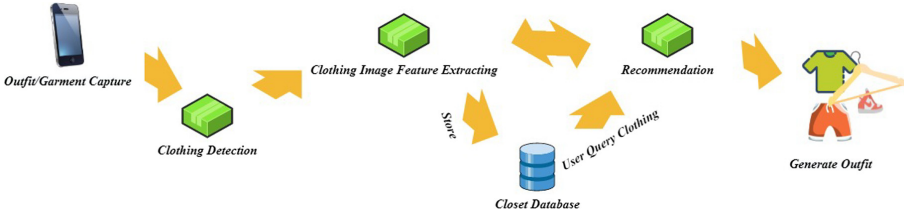

Fig. 2. Workflow for extracting the clothing attributes and recommendation.

The image of the clothing can be obtained through a photograph or an online store, where it is processed through Convolutional Neural network (CNN) to classify and extract the characteristics of the garment. Subsequently, its characteristics will be saved in the database to be used during the combination task. This database represents the virtual wardrobe of the blind user.

The high-level steps associated to the clothing combination are described as:

- Clothing Detection: recognize/segment the garment contained in the photography and assign it to the categories upper, lower, shoes or accessories.
- Clothing Image Feature Extraction: extract features/attributes of the garment as: type, season, pattern.
- Recommendation: suggest a combination for a garment, fulfil an incomplete outfit and suggest a complete outfit.

Neural networks will be used for clothing classification and combination. For the extraction of characteristics, CNNs will be used due to the good results demonstrated in the state of the art.

During a process of combinations, the algorithm will learn the user's preferences based in its frequency of use.

As the target group is blind people and in order to provide good usability practice, a friendly interface will be created, providing mechanisms to assure the proper accessibility by smartphone or tactile screen. A database will be designed to record all data processed and acquired, including the results of statistical analysis. The software will be designed with multiplatform compatibility, in the sense of being later integrated into several operating systems.

Later, a mechatronic system with robotics will be designed in order to integrate all the developed software. This system will make the entire process of choosing, identifying and selecting garments autonomously.

Finally, test trials are essential to assess the proper functioning of the system. Several tests will be considered involving a sufficient number of blind people, as well as worst-case scenarios to test the robustness of the algorithms for real time image processing and data acquisition. This task must act in an iterative approach with the other tasks, in order to allow the optimization of the system, since the collection of quantitative data will be essential until obtaining the final version.

5 Experiments

In this work we decided to build a dataset in order to identify how its construction can influence the results.

The dataset was based on images collected from the internet and distributed in 7 categories, according to the Table 1. In online shops and e-commerce stores the photos have a white background with only one item of garment by each photo, that is exactly the opposite of the pictures in our dataset where most of them have a background and more that one part of the body.

Table 1. Dataset description

Category	Number of samples
Ankle boot	105
Bag	105
Coat	105
Pullover	105
Sneaker	105
T-Shirt	105
Trouser	105

In order to achieve good results, it is essential to take into consideration image processing before introducing the images in neuronal networks. However, we try to approximate our dataset to possible real-life cases that could occur with the user taking a photo of his/her clothe. In addition, image processing requests more resources and time.

Based in the success of the works mentioned in the state of the art we adopted in our approach the convolutional neuronal networks for feature extracting and image classification.

In order to evaluate the data and starting sketch our algorithm we choose two types of CNN applied in ImageNet Large Scale Visual Recognition Challenge [22]. The goal of ImageNet is object detection and image classification where an input image is classified in one of the 1,000 categories.

In this context, we have implemented the VGG16 from scratch in Keras due to building simplicity using 3×3 convolution layers with a softamx classifier after the fully connected layers as show in Fig. 3.VGG means Visual Geometry Group at University of Oxford and the number "16" represents the weight layers in the network.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64	conv3-64	conv3-64	conv3-64	conv3-64
	LRN	**conv3-64**	conv3-64	conv3-64	conv3-64
maxpool					
conv3-128	conv3-128	conv3-128	conv3-128	conv3-128	conv3-128
		conv3-128	conv3-128	conv3-128	conv3-128
maxpool					
conv3-256	conv3-256	conv3-256	conv3-256	conv3-256	conv3-256
conv3-256	conv3-256	conv3-256	conv3-256	conv3-256	conv3-256
			conv1-256	**conv3-256**	conv3-256
					conv3-256
maxpool					
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
			conv1-512	**conv3-512**	conv3-512
					conv3-512
maxpool					
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
			conv1-512	**conv3-512**	conv3-512
					conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Fig. 3. ConvNet configurations [16]

The dataset was split in 80% for training and 20% for testing.

The model achieved 80% accuracy in the training set and 66% in the testing set.

Figure 4 shows the behaviour relative to the accuracy and loss during the training and testing process. The graph shows that the best accuracy was achieved with 30 epocs, while obtaining the lowest loss value in the model.
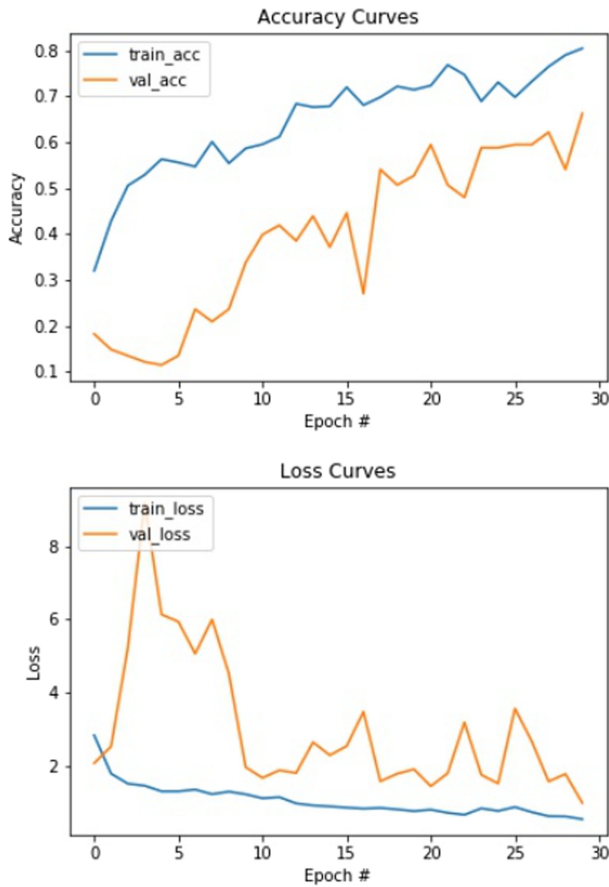
Accuracy Curves

Loss Curves

Fig. 4. Loss and accuracy of VGG16 model (training process in blue and testing process in orange).

After the training and validation phases, some images were submitted to the model for classification. Examples of classification can be verified in Fig. 5. Although in Fig. 5 are shown good classification scores, it was also observed some misclassification in particular, in garments with similar shapes.

A recent Google study shows that training from fine tuning is better than training form scratch since it is required less cost in terms of data and time to training [23]. In this sense, a ResNet50 [24] pretrained ImageNet weights was used. It was replaced the top of original architecture, the fully-connected layers and softmax classifier by new ones in order to learn the new classes. In this way only a head part of the network is re-trained.

Fig. 5. Example of garment classification.

The fine-tuning of ResNet50 was trained under Google Colab that is a free cloud service hosted by Google. The major advantage of this notebook is that it provides a free GPU [25].

The results obtained with this fine-tuning training are approximated to the previous network where 63% accuracy in training set and 71% in accuracy test set were achieved. Figure 6 shows the behaviour relative to the accuracy and loss during the training and testing process in ResNet50. It is possible to observe that during the build of the model, accuracy and loss in test set have achieved a better result than in the training set.

Comparing the graphs of the training and validation accuracy and loss from both configurations, it is possible to see that the ResNet50 needs more epochs to achieve the same accuracy, Fig. 6. Note that in the VGG16 with 30 epochs it is achieved the best performance possible. Otherwise the last one needs 60 epochs to achieve the same results.
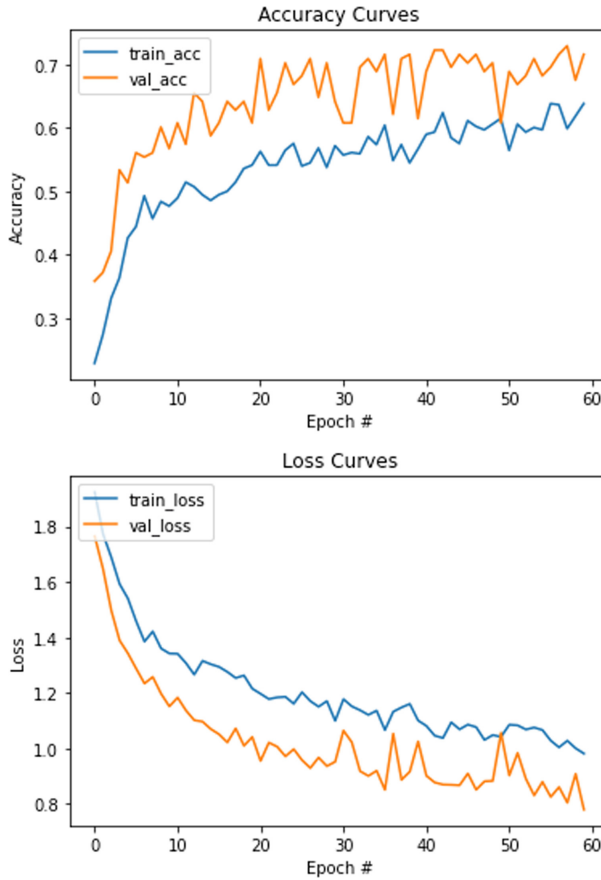
Fig. 6. Accuracy and loss in ResNet50 (training process in blue and testing process in orange).

6   Final Remarks

It is recognized that the aesthetics of clothing is an important issue in human lives.

Blind people may face additional problems in accomplishing this issue, namely in identifying and combining clothing. Some are helped by family or friends, others need a pronounced organizational capacity. This system proposes a new concept to significantly improve the daily life of blind people, allowing the blind to identify and combine clothing pieces.

During this paper we have explored two different convolutional neuronal networks being one of them implemented from scratch and the other one through fine-tuning.

The amount of data on each category is not enough to reach good results. The deep learning is adequate for big amount of data, meaning that it is necessary to gather more data in order to improve the results.

The results obtained by using a simple network designed from scratch and with a fine-tuning network were similar. Moreover, to process the fine tuning in ResNet50 has higher computational and time costs.

In the future, we will experiment large datasets, extracting and identifying new attributes. Some additional image processing, like remove background and body parts will be performed in order to improve the results.

Finally, we intend to implement our algorithm on a web platform or mobile device, in order to start testing interactivity with blind users. After that, the smart wardrobe will also be built to incorporate all the developed algorithms.

References

1. Rocha, D., Carvalho, V., Soares, F., Oliveira, E.: Design of a smart mechatronic system to combine garments for blind people: first insights. In: Garcia, N.M., Pires, I.M., Goleva, R. (eds.) HealthyIoT 2019. LNICSSITE, vol. 314, pp. 52–63. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-42029-1_4
2. Rocha, D., Carvalho, V., Soares, F., Oliveira, E.: Extracting clothing features for blind people using image processing and machine learning techniques: first insights. In: Tavares, J.M.R.S., Natal Jorge, R.M. (eds.) VipIMAGE 2019. VipIMAGE 2019. Lecture Notes in Computational Vision and Biomechanics, vol. 34, pp. 411–418. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32040-9_42
3. Stylebook Closet App: a closet and wardrobe fashion app for the iPhone, iPad and iPod. http://www.stylebookapp.com/index.html. Accessed 21 June 2019
4. ShopStyle: Search and find the latest in fashion. https://www.shopstyle.com/. Accessed 21 June 2019
5. Tailor : The smart closet. http://www.tailortags.com/. Accessed 21 June 2019
6. Colorino Color Identifier - Light Detector - Assistive Technology at Easter Seals Crossroads. https://www.eastersealstech.com/2016/07/05/colorinos-color-identifier-light-detector/. Accessed 19 Feb 2019
7. Colortest Standard - Computer Room Services. https://www.comproom.co.uk/product/colortest-classic/. Accessed 19 Feb 2019
8. Yamaguchi, K., Kiapour, M.H., Ortiz, L.E., Berg, T.L: Parsing clothing in fashion photographs. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3570–3577 (2012)
9. Yamaguchi, K., Kiapour, M.H., Ortiz, L.E., Berg, T.L.: Retrieving similar styles to parse clothing. IEEE Trans. Pattern. Anal. Mach. Intell. **37**, 1028–1040 (2015). https://doi.org/10.1109/TPAMI.2014.2353624
10. Wazarkar, S., Keshavamurthy, B.N.: Fashion image classification using matching points with linear convolution. Multimedia Tools Appl. **77**(19), 25941–25958 (2018). https://doi.org/10.1007/s11042-018-5829-4
11. Yang, M., Yu, K.: Real-time clothing recognition in surveillance videos. In: 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 2937–2940 (2011)

12. Yuan, S., Tian, Y., Arditi, A.: Clothing matching for visually impaired persons. Technol. Disabil. **23**, 75–85 (2011). https://doi.org/10.3233/TAD-2011-0313

13. Yang, X., Yuan, S., Tian, Y.: Assistive clothing pattern recognition for visually impaired people. IEEE Trans Hum-Mach. Syst. **44**, 234–243 (2014). https://doi.org/10.1109/THMS.2014.2302814

14. Wang, W., Xu, Y., Shen, J., Zhu, S.-C.: Attentive fashion grammar network for fashion landmark detection and clothing category classification. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 4271–4280 (2018)

15. Liu, Z., et al.: DeepFashion: powering robust clothes recognition and retrieval with rich annotations. In: 2016 Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1096–1104 (2016)

16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR 2015) - Conference Track Proceedings (2015)

17. Cychnerski, J., et al.: Clothes detection and classification using convolutional neural networks. In: 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8 (2017)

18. Hara, K., Jagadeesh, V., Piramuthu, R.: Fashion apparel detection: the role of deep convolutional neural network and pose-dependent priors. (2014). https://doi.org/10.1109/WACV.2016.7477611

19. Rocha, D., Carvalho, V., Gonçalves, J., Azevedo, F., Oliveira, E.: Development of an automatic combination system of clothing parts for blind people: MyEyes. Sensors Transd. **219**(1), 26–33 (2018)

20. Rocha, D., Carvalho, V., Oliveira, E., Gonçalves, J., Azevedo, F.: MyEyes-automatic combination system of clothing parts to blind people: first insights. In: 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), pp 1–5. IEEE (2017)

21. Rocha, D., Carvalho, V., Oliveira, E., Gonçalves, J., Azevedo, F.: MyEyes-automatic combination system of clothing parts to blind people: first insights. In: 2017 IEEE 5th International Conference on Serious Games and Applications (SENSORDEVICES 2017), Italy, pp. 11–14 September 2017

22. ImageNet Large Scale Visual Recognition Competition (ILSVRC). http://www.image-net.org/challenges/LSVRC/. Accessed 13 July 2020

23. Kornblith, S., Shlens, J., Le, Q.V.: Do better ImageNet models transfer better?. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2656–2666 (2018). https://doi.org/10.1109/CVPR.2019.00277

24. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition, In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90

25. Welcome To Colaboratory - Colaboratory. https://colab.research.google.com/notebooks/intro.ipynb. Accessed 14 July 2020