Enhancing STEM Education using Augmented Reality and Machine Learning

Ivan Jie Xiong Ang, King Hann Lim
Department of Electrical and Computer Engineering
Curtin University, Malaysia
CDT 250, 98009, Miri, Sarawak, Malaysia
Email: ivanangjx@gmail.com

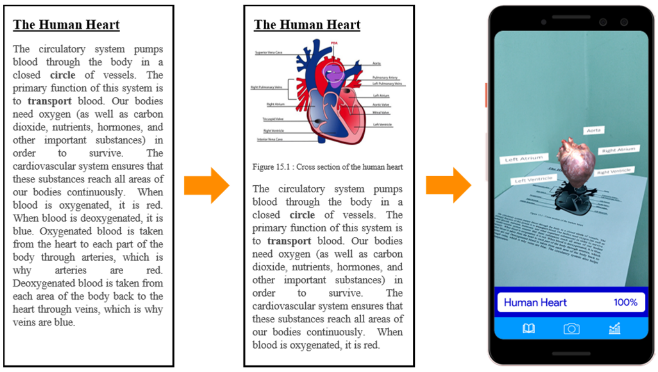
Abstract*—Learning Science, Technology, Engineering and Mathematics (STEM) in the 21st century has been evolved from the conventional textbook to the interactive platform using electronic devices. This paper presents the implementation of a mobile application system, named AUREL (Augmented Reality Learning) in enhancing the learning experience by projecting Augmented Reality (AR) objects onto 2D images. This AR visualization is used to improve the understanding of STEM subjects and increases the enthusiasm of students towards STEM subjects. In this implementation, Google's Cloud Tensor Processing Units (TPUs) are used to train specific datasets alongside Cloud Vision API to detect a wide range of objects. ML Kit for Firebase is used to host the custom TensorFlow Lite models for specific use cases for better accuracy. On the other hand, Google Cloud Platform (GCP) is used to harvest STEM data, manage STEM 3D information and data processing. Subsequently, the processed information will be displayed in AR in the mobile application using ARCore's Sceneform SDK. The application of AUREL could be extended to all science subjects so that students can learn using an interactive platform.

***Keywords*— Augmented Reality, Machine Learning, STEM Education**

I. Introduction

The notion of science, technology, engineering and mathematics (STEM) is an integrated from of curriculum, information and assessment to reinforce science education by seamlessly combining these four disciplines [1]. STEM education plays a key role in the sustained growth and stability of Malaysia's economy by creating critical thinkers and the next generation of innovators [2]. Most careers in the future will require a basic understanding of science and math. Despite these compelling facts, the STEM enrolment in Malaysia's education system has dropped below 20% as of 2017. This may lead to insufficient number of STEM graduates which will derail the nation's development plan. Hence, this project is addressed to improve the enthusiasm and student's attitude towards STEM by providing an alternate way for students to learn.

The conventional teaching method using textbooks has been adopted since 19th century to deliver knowledge. Teachers and students rely heavily on textbooks to pass and receive information [3]. Textbooks often contains paragraphs of texts for students to read from as seen in Fig. 1(a). Sometimes, textbooks will include pictures and diagrams to better illustrate the information so that students can get a better understanding as seen in Fig. 1(b). However, in many cases, pictures are unable to convey the necessary information needed to fully grasp the knowledge of the subject matter. For example, it is hard to visualize the effects of forces, the vibrations of atoms or internal organs of the human body. Nowadays, smartphone devices are common



(a) (b) (c)

Fig. 1:(a) Paragraph in textbooks (b) Illustration and paragraph in textbooks (c) Augmented Reality in AUREL

gadgets among students. However, these electronic devices are more commonly used for browsing social media and playing games.

Therefore, this paper proposes the creation of a mobile application called AUREL (short for Augmented Reality Learning) to make use of these mobile devices more beneficially, which is in the student's education. E-learning, which is the learning conducted via electronic media, has emerged as a promosing method to replace conventional learning methods [4]. AUREL aims to be an E-learning tool for students to better understand existing textbooks by combining the capabilities of Machine Learning and Augmented Reality (AR). AUREL also addresses the ability to accurately perform image labelling on a wide array of objects in a resource constrained environment such as mobile devices. Furthermore, AUREL explores the usage of markerless AR in education so that the display of AR objects are not limited to a specific image [5].

II. Overview of AUREL System

AUREL makes full use of Google's Cloud Platform services to be able to function both online and offline. While connected to the Internet, Cloud Vision API is used for image labelling as shown in Fig. 2(a). Vision API quickly classifies images into thousands of categories and improves over time as new Machine Learning concepts are introduced. A MobileNets model will be used for on-device labeling instead if there is no Internet connection as seen in Fig. 2(b). ML Kit for Firebase enables the application of Vision API onto the mobile application as well as hosting custom TensorFlow Lite models. While online, these TensorFlow Lite models are used to label categories in a more detailed manner at which the Cloud Vision API is unable to. These models are also automatically downloaded so that on-device labelling can be done in offline situations as seen in Fig. 2(b). Specific

978-1-7281-1557-3/19/$31.00 ©2019 IEEE

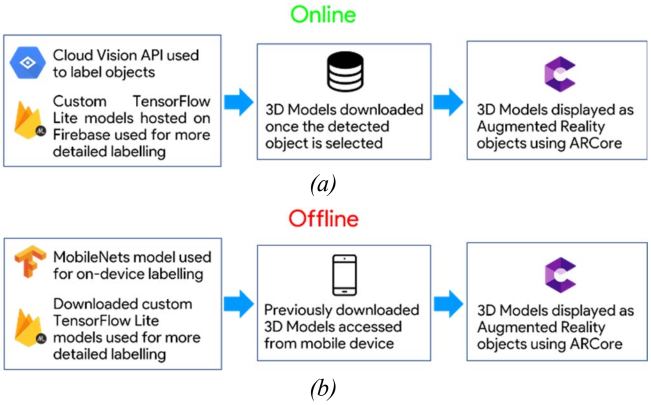


Fig. 2: Overview of AUREL system during (a) online and (b) offline situations

datasets are trained on Cloud TPUs to obtain more accurate labels for these specific images. Training machine learning models on Cloud TPUs, which is computationally demanding, eliminates the need for expensive and powerful hardware locally.

Once the user confirms the selection, the 3D model of the object will be downloaded from an online repository onto the device as shown in Fig. 2(a). The downloaded 3D models can be accessed by the user even in offline situations as seen in Fig. 2(b). Augmented Reality is a technique of showing virtual objects on real-world images which is enabled by ARCore [6]. Once the 3D model of the selected object is downloaded or accessed from the mobile device, ARCore and Sceneform are used to project the markerless Augmented Reality object as seen in Fig. 1(c). The user can interact with the object and have a much more immersive learning experience.

III. MobileNets Architecture

In AUREL, deep learning is implemented to enable the detection of objects in an unconstrained environment. Network architectures such as AlexNet [7], Inception-V3 [8] and MobileNets [9] are popular for mobile applications. A balance between resource constraints, speed and time is important in the miniature nature of mobile devices [10]. AUREL uses MobileNets for image classification due to its strength in minimizing time and space for image classification while only compromising the accuracy slightly [9]. The architecture of MobileNets uses depthwise separable convolutions (Conv dw) instead of commonly used convolutional layers as seen in Fig. 3. The depthwise separable convolution block is split into two sections: first the depthwise convolution layer filters the input, then the pointwise convolution layer combines these filtered values to create new features. It performs the same as traditional convolution but is considerably faster.

The full architecture of MobileNets consists of a 3×3 convolution as the first layer, followed by 13 times the depthwise separable convolution block as seen in Table I. With an input image of 224×224×3, the output of the network will be a 7×7×1024 feature map. The implementation of depth wise separable convolution significantly reduces the number of parameters compared to networks with normal convolutions with the same depth in the networks. This causes the reduction in total number of floating point

TABLE I: MobileNets Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ *dw*	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ *dw*	$112 \times 112 \times 64$
Conv s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ *dw*	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ *dw*	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ *dw*	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ *dw*	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5 × Conv dw / s1	$3 \times 3 \times 512$ *dw*	$14 \times 14 \times 512$
5 × Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ *dw*	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ *dw*	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×10	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 10$

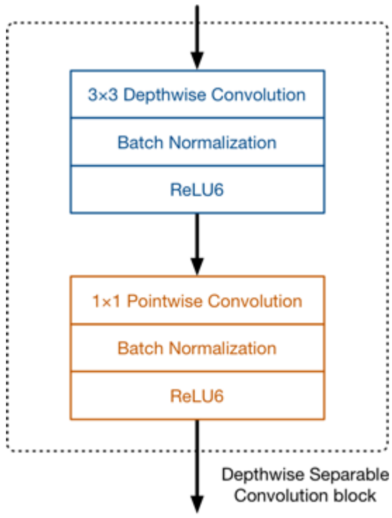


Fig. 3: Depthwise Separable Convolution Block [11]

multiplication operations which is favorable in mobile and embedded applications with less compute power. Therefore, MobileNets is chosen for this application because it can perform much faster than comparable neural networks with just some sacrifice of accuracy.

IV. Implementation of Google Cloud Platform

A. Cloud Vision API & Custom TensorFlow Lite Models

AUREL not only enables users to browse for a broad variety of subjects, but also objects within a specific theme or category. AUREL enables users to select between using Cloud Vision API for Broad Learning or custom TensorFlow Lite Models for Themed Learning to detect objects. In Broad Learning, Google Cloud Vision API is used for image classification. The Cloud Vision API is trained on a large dataset of images and can provide 10,000+ labels in many categories while the mobile application is connected to the cloud. In situations where there is no Internet connection, the on-device API is still able to provide 400+ labels that cover the most commonly found concepts in images.

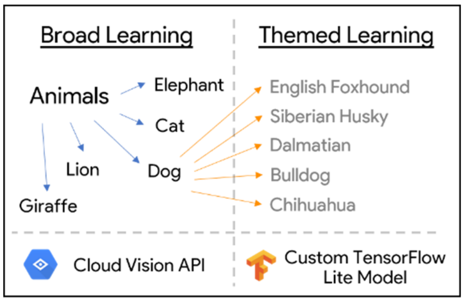


Fig. 4: Use cases for Broad Learning and Themed Learning

Images	Cloud Vision API		Custom TensorFlow Lite Model	
	Dog	97%	Golden Retriever	99%
	Canidae	94%	English Retriever	83%
	Dog Breed	93%	Norwich Terrier	55%
	Dog	98%	Siberian Husky	98%
	Canidae	94%	Alaskan Husky	94%
	Dog Breed	93%	Akita	86%
	Dog	95%	English Springer	95%
	Canidae	94%	Welsh Springer	94%
	Dog Breed	92%	English Setter	92%

Fig. 5: Comparison results between Cloud Vision API and Custom Tensorflow Lite Model

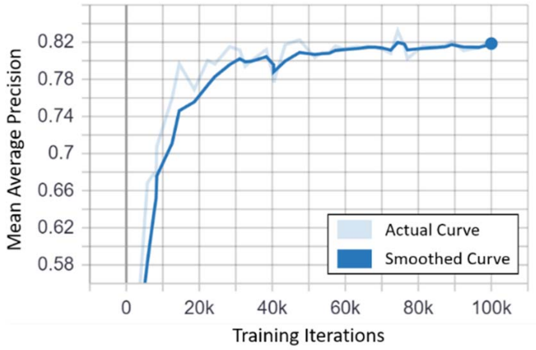
For example, in Fig. 4, if the user wishes to learn about animals in general, Broad Learning is used. However, if the user wishes to go one step further and identify the species of the dog, then Themed Learning is selected whereby the custom TensorFlow Lite model is used. As accurate as the Cloud Vision API may be in identifying a wide range of objects, it is not feasible in this situation. As seen in Fig. 5, Cloud Vision API is only able to label the images as dogs. A TensorFlow Lite model specifically trained on different species of dogs can be used to accurately identify the species of the dog. The dog species Golden Retriever, Siberian Huskey and English Springer were correctly identified as seen in Fig. 5.

B. Obtaining & Preparing Dataset

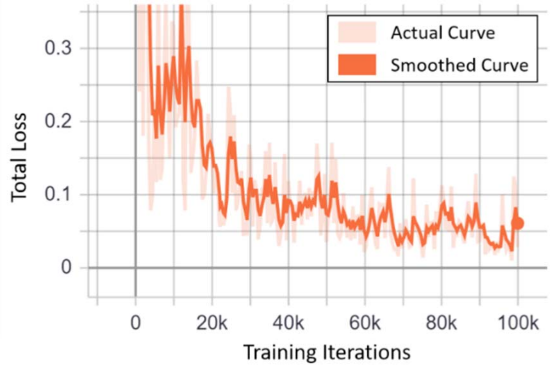
Datasets for STEM learning must be prepared for network training. Google Images Download is an open source tool that can be used for searching and downloading Google images. The tool can be configured to search for keywords, file types, size and color filters of images. Each themed dataset which consists of 10 classes is used to demonstrate this project. Each class consists of 100 images that will be used for training.

C. Training the Model

AUREL aims to enable users to perform on-device image classification in offline situations. Therefore, models must be able to run quickly with high accuracy in a resource-constrained environment making use of limited computation, power and space [12]. In this project, quantization-aware training is done on MobileNets. We chose MobileNets because it is a small and efficient convolutional neural network which is designed to efficiently maximize accuracy while considering the restrictive resources for embedded applications [13]. The selected MobileNets model is trained using transfer learning on the ImageNet Large Visual Recognition Challenge Dataset [14]. The MobileNets model



(a)



(b)

Fig. 6: (a) Accuracy of model over training iterations, (b) Loss of model over training iterations

contains networks with millions of parameters that can differentiate many classes. By using these parameters as inputs to the final classification layer, a model can be trained accurately.

Tensorboard is used to monitor and inspect the training as it takes place. As the training progresses, a series of output steps showing the training and validation accuracy is displayed on Tensorboard as shown in Fig. 6. The training can be set to end at a predefined number of steps or until an accuracy percentage is exceeded. For 10 classes each having 100 images, the batch size equals to 1k. The training is set to stop upon reaching 100k steps (100 epochs) or until an accuracy of 90% is obtained. As seen in Fig. 6(a), the training stops at 100k steps since the preset accuracy of 90% was not met. The total loss at the end of training is 0.0532 as seen in Fig. 6(b).

D. Converting the model to TensorFlow Lite format

TensorFlow Lite is TensorFlow's lightweight solution for mobile applications and embedded devices by enabling on-device machine learning inferences with low latency and a small binary size [15]. TensorFlow Lite Optimizing Converter is used to optimize the trained TensorFlow model. Post-training quantization is performed to reduce the model size, at the same time providing up to 3 times lower latency with only a fraction of degradation in model accuracy. The converter also prunes unused graph-nodes and improves performance by joining operations into more efficient composite operations. While optimizing TensorFlow protobuf files which contains graph definitions and weights of the model, TensorFlow Lite uses a different serialization format which is FlatBuffers. FlatBuffers can be memory-

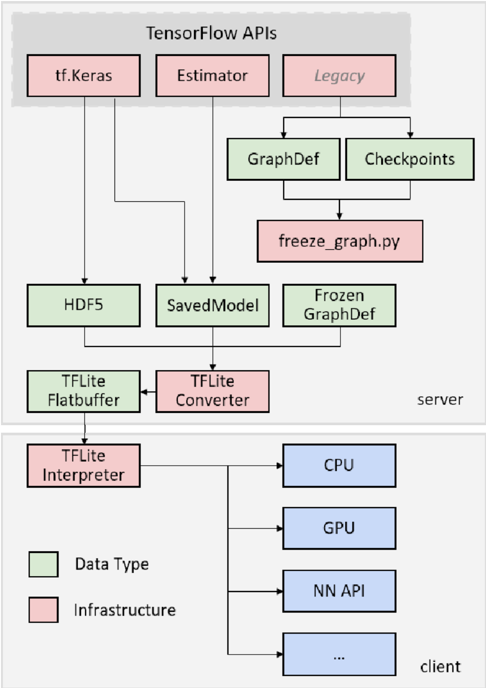


Fig. 7: TensorFlow Lite conversion process

mapped and used directly from disk without having to be loaded and parsed which allows faster startup times. Fig. 7 depicts the TensorFlow Lite conversion process.

E. Hosting TensorFlow Lite Models on ML Kit for Firebase

ML Kit is a mobile Software Development Kit (SDK) which eases the implementation of machine learning into mobile applications. ML Kit brings together Google Cloud Vision API, Mobile Vision and TensorFlow Lite together into a single SDK. There are many benefits of hosting TensorFlow Lite models on the cloud instead of on-device. Model versions can be easily managed and updated and will significantly reduce the file size of the AUREL application. The image classification models will only be downloaded onto AUREL if requested by the user.

F. Storing the 3D models on Google Cloud Storage

3D models normally consists of 2 files, the geometry definition of the object (.obj) and the textures applied to the object (.mtl). These 2 files of an object are used to create a Sceneform Asset file (.sfb) which is used by the mobile application to create a renderable object to be displayed in Augmented Reality. A repository of these Sceneform Asset files is stored on Google Cloud Storage. Each asset has its unique directory and will be downloaded onto the mobile application once the user has identified an object and wants to display it in Augmented Reality.

V. AUREL Mobile Application

AUREL is developed on Android Studio 3.0 and it can be installed on Android devices with minimum Android 4.0 (Ice Cream Sandwich). The ARCore functionality of AUREL requires devices with minimum Android 7.0 (Nougat). Throughout the development of AUREL, Android Emulator is used to test the app to ensure that backend systems and user interface functions properly with different versions of Android and screen resolutions. The initial launching of

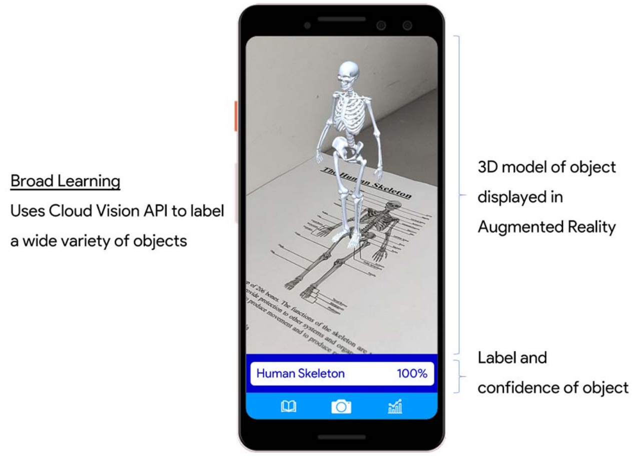


Fig. 8: Broad Learning page

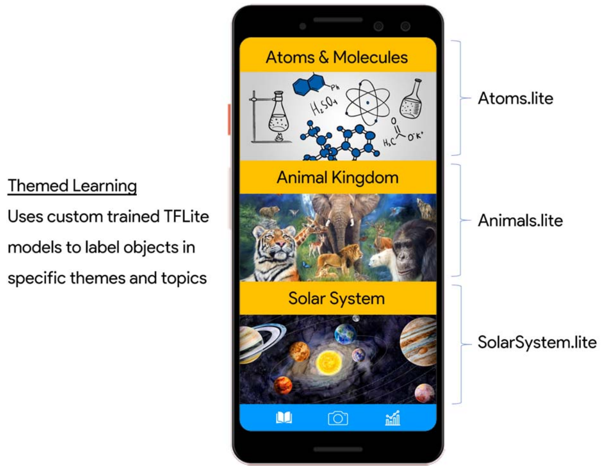


Fig. 9: Themed Learning page

AUREL system requires the user confirmation to enable access to the device's camera. Upon accepting, the user will be directed to the main page of the application. There are three main pages on the application which are Broad Learning, Themed Learning and History and analytics.

In Broad Learning, the user can take a picture of any object that the user is interested to learn more about. If the device is connected to the Internet, the application uses Cloud Vision API to detect the object in the picture. This functionality allows the user to search for objects even if they don't know what the name of the object is. A list of the top results and the corresponding confidence is displayed on the screen of the device. Once the user selects one of the results, its 3D model will be downloaded from the 3D model repository on Google Cloud Storage onto the device. The user is then able to point their device onto a flat surface and display the 3D model of the object in Augmented Reality using ARCore. The user can interact with the object such as rotating and scaling the object to fully visualize the object in real life. In offline situations, general learning functions are the same, the only downside is that image labeling will be slightly less accurate as the on-Device API is used instead.

Less common objects would not be detected accurately by the application. In addition, new 3D models would not be able to be downloaded onto the device directly. The application will store the data and will only download the 3D model once the device is connected to the Internet. It is strongly suggested

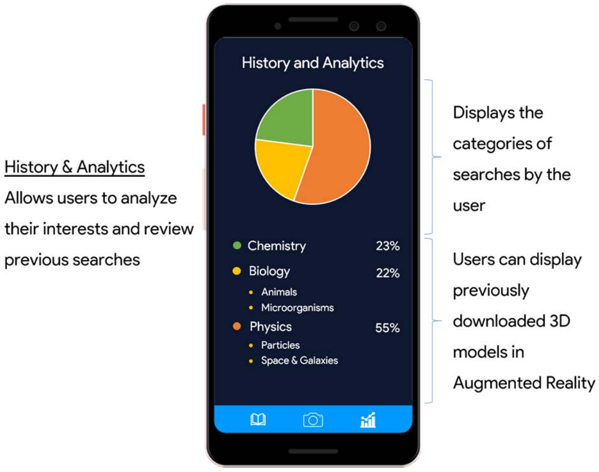


Fig. 10: History & Analytics page

that users have at least basic Internet connection while using the application to obtain optimal results and to fully utilize the functionalities.

Themed Learning is used for situations whereby the user is interested to explore specific topics such as "Atoms and Molecules", "Animal Kingdom" and "The Galaxy". Each of these topics will use their respective TensorFlow Lite models to identify objects for the user. This allows more in-depth labelling which the Cloud Vision API might not be able to provide with the required accuracy. These TensorFlow Lite models are hosted on MLKit for Firebase and will only be downloaded onto the application when required hence decreasing the file size of the application. This also allows the Image Labelling models to be constantly updated on the cloud without the user having to reinstall the mobile application each time the models are updated.

All previously downloaded 3D models are stored on the device so that the user can be able to view the models in Augmented Reality anytime, anywhere. This enables the user to learn at any given moment such as while waiting at the bus stop. The search history of the user is also properly documented into categories such as Physics, Biology, Astronomy and many more. These data are displayed statistically so that the user can get a better understanding of their interests. Another feature is that the user can send feedback on wrongly identified objects so that the machine learning models can be corrected.

VI. Conclusion & Future Improvement

Conventional teaching and learning methods has not been effective in educating the next generation of STEM graduates. A mobile platform combining Augmented Reality and Machine Learning, which is named as AUREL is proposed in this paper to improve the STEM education. This platform leverages with the capability of Google Cloud Platform to perform machine learning on markerless object to retrieve useful information for 3D visualization. Broad learning and themed learning is added in the AUREL to ignite students' passion and curiosity towards specific STEM subjects. For future works, 3D models that includes animations can be added to gain a better understanding of the object in concern. Mixed reality and Virtual Reality can also be used alongside this application to improve the impressiveness of the learning experience for the users. Users can be transported into another dimension and be fully immersed in learning new discoveries. Future versions of this application will include the ability to automatically update the Image Labelling models by using the images taken by the users as inputs to improve the model.

References

[1] Breiner, J.M., Harkness, S.S., Johnson, C.C. and Koehler, C.M., "What is STEM? A discussion about conceptions of STEM in education and partnerships.", School Science and Mathematics, 112(1), pp.3-11, 2012.

[2] Yu Xie, Michael Fang and K. Shauman, "STEM Education", in Annual Review of Sociology Vol. 41:331-357, 2015.

[3] Wei Li, Yap, "Transforming Conventional Teaching Classroom to Learner-Centred Teaching Classroom Using Multimedia-Mediated Learning Module", in International Journal of Information and Education Technology, no. 6. 105-112, 2016.

[4] Dongsong Zhang, J. Leon Zhao, Lina Zhou, and Jr., Jay F. Nunamaker. "Can E-Learning Replace Classroom Learning?" Communications of the ACM 47, no. 5, 2004.

[5] Yuen, S.C.Y., Yaoyuneyong, G. and Johnson, E., "Augmented reality: An overview and five directions for AR in education". Journal of Educational Technology Development and Exchange (JETDE), 4(1), p.11, 2011.

[6] Güngör, Cengiz, "Augmented Reality Development Tools and Google ARCore", 2017 1st International Symposium on Multidisciplinary Studies and Innovative Technologies, At Tokat, Turkey, 2017.

[7] Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Van Esesn, B.C., Awwal, A.A.S. and Asari, V.K., "The history began from AlexNet: a comprehensive survey on deep learning approaches." arXiv preprint arXiv:1803.01164, 2018.

[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 2818-2826, 2018.

[9] N. R. Gavai, Y. A. Jakhade, S. A. Tribhuvan and R. Bhattad, "MobileNets for flower classification using TensorFlow," *2017 International Conference on Big Data, IoT and Data Science (BID)*, Pune, 2017, pp. 154-158, 2017.

[10] J. Wang, B. Cao, P. Yu, L. Sun, W. Bao and X. Zhu, "Deep Learning towards Mobile Applications," *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, Vienna, 2018, pp. 1385-1393, 2018.

[11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251-1258, 2017.

[12] N. D. Lane, S. Bhattacharya, A. Mathur, P. Georgiev, C. Forlivesi and F. Kawsar, "Squeezing Deep Learning into Mobile and Embedded Devices," in *IEEE Pervasive Computing*, vol. 16, no. 3, pp. 82-88, 2017.

[13] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861, 2017.

[14] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255, 2009.

[15] Ignatov, Andrey, Timofte, Radu, Szczepaniak, Przemyslaw, Chou, William, Wang, Ke, Wu, Max, Hartley, Tim and Van Gool, Luc. "AI Benchmark: Running Deep Neural Networks on Android Smartphones", 2018.