EAI Endorsed Transactions
on Creative Technologies
Research Article EAI.EU

TailorEd: Classroom Configuration and Activity Identifiers (CCID & CAID)

Andres Calle[1], Quan Nguyen[1], Kristin Lee[3], Julia Voss[2], and Navid Shaghaghi[1,3,4,*]

Ethical, Pragmatic, and Intelligent Computing (EPIC) Research Laboratory
[1]Department of Computer Science and Engineering (CSEN)
[2]Department of English
[3]Department of Information Systems and Analytics (ISA)
[4]Department of Mathematics and Computer Science (MCS)
Santa Clara University (SCU), Santa Clara, California, USA

Abstract

INTRODUCTION: The study of how classroom layout and activities affect learning outcomes of students with different demographics is difficult because it is hard to gather accurate information on the minute by minute progression of every class in a course. Furthermore, the process of data gathering must produce an abundance of data to work with and hence must be automated.
OBJECTIVES: A machine learning model trained on images of a classroom and thus capable of accurately labeling the classroom layout and activity of many thousands of images much faster and cheaper than employing a human.
METHODS: Transfer learning can allow for preexisting computer vision models to be retrained on a smaller, more specific dataset in order to still achieve a highly accurate result.
RESULTS: In the case of the classroom layout, the final model achieved an accuracy of 97% on a four category classification. And for detecting the classroom activity, after experimentation with several different versions that could work on a very small sample sizes, the best model achieved an accuracy of 86.17%.
CONCLUSION: In addition to showing that using computer vision to determine human activities is possible albeit more difficult than layouts of inanimate objects such as classroom desks, the study shows the differences between the use of self-supervised learning techniques and data augmentation techniques in order to overcome the problem of small training data-sets.

Received on 11 December 2021; accepted on 06 April 2022; published on 28 July 2022

Keywords: Classroom Configuration, Convolutional Neural Networks (CNNs), Data Science, Education Technology, Image Recognition

doi:10.4108/eetct.v9i3.2229

1. Introduction

The study of how well different student demographics perform under different classroom environments has always been a source of interest in pedagogical research. However, two of the limiting factors of studies into classroom environments has been the trade-off between small sample sizes or and the use of self-reported data. Researchers simply can not be placed in every class to take minute by minute observations on the layout of the classroom or type of activity being performed by the class. The other alternative, however, is to rely on inaccurate self-reported data from teachers and students on how much time they spent on each activity.

Advances in machine learning and computer vision have created a solution to this problem: any classroom containing an installed camera can have a minute by minute report on the classroom environment and class activities. This novel solution would allow simply for the installation of a camera within a classroom in order to obtain accurate data on the class, allowing the instructor to have a real time breakdown on

*Corresponding author. Email: nshaghaghi@scu.edu

how class time and environment is being used as well as the expected influence on learning outcomes for different groups of students that teaching style would produce. TailorEd - an automated system under research and development at Santa Clara University's Ethical, Pragmatic, and Intelligent Computing (EPIC) research laboratory - utilizes aerial classroom photos (such as seen in Figure 1.6) taken at 1-minute intervals from courses taught in 2014-2016, to study the relationship between classroom design and activities with student learning.

Figure 1. Example Classroom Image

By providing the images taken by the camera to a trained machine learning image classifier, both the physical layout of the classroom and the activity that the classroom occupants are engaging in can be determined quickly. The development of convolutional neural networks such as ResNet has resulted in computer vision algorithms with recognition capabilities on par with human vision, and through the use of Transfer Learning the models can be accurately retrained to accomplish new tasks such as determining the activity type of a classroom. With this predictive model, studies can now avoid the pitfalls of drawing conclusions from small datasets or faulty self-reported data and take advantage of the automated labeling of large datasets of photos. This paper details how the aforementioned automated machine learning classifier is built, trained, and validated. In the Background section we will cover what the parameters of the tasks required for the machine learning algorithms are as well as why they are being performed. The Methodology section will cover the two specific models used, how the are constructed, what function they perform, what output would be expected from them. The Results section goes over the experiments performed to construct the most accurate version of the two models possible, as well as examining why some results are worse than others. Finally the future works and conclusion sections wrap up the results of this paper as well as cover future areas of expansion for this research.

2. Background

The TailorEd system currently supports two models for classroom analysis using classroom photos to determine the layout of the classroom and the pedagogical activity being performed in the classroom at the time the images are taken.

- The Classroom Configuration Identifier (CCID) uses an image recognition Convolutional Neural Network (CNN) known as AlexNet in order to determine whether the layout of the classroom is 1. Forward-Facing; 2. Circular or 'U' Shaped; 3. Small Groups; or 4. Empty. These layouts refer to the physical space the students are occupying, agnostic of any particular activity that is being performed by them. As such, classrooms with fixed desks will only exist in two configurations: filled or empty, while classrooms with movable desks can change configurations many times over even the course of a single class session.
- The Classroom Activity Identifier (CAID) which uses a different CNN known as ResNet is focused on determining the specific activity the students are engaged in the classroom. It can distinguish between six different activity types (1. Empty; 2. Lecture; 3. Discussion; 4. Group Work; 5. Reading; 6. Writing) and classification is performed agnostic of any particular classroom layout.This particular model can thus be used effectively on both traditional classrooms and classrooms with movable desks (unlike the CCID model).

2.1. Why study classroom configuration

According to a study performed by Wannarka and Ruhl, desk arrangement can have a substantial effect on students' behavior and overall performance in a course. They explain that factors that influence communication such as orientation and proximity reveal themselves in desk formations, and consequently contribute to the quality and extent of student interaction [5]. Depending on how much interaction is desired throughout a course, this configuration is something that teachers must take into account. Unfortunately, there is currently limited knowledge available for making these decisions effectively and readily. While teachers may learn from past experiences which formations are conducive and which pose a hindrance to the learning process, due to the ever-changing class composition, what works for one group of students may not work for others. Furthermore, what may work for one subject may not work for another. It stands to reason that the pedagogies typical of philosophy require a lot of group discussion and argumentation, which differ from those typical of

computer science, which require hands on experience with individual and peer coding, which differ from those of teaching sculpture using tools not normally found in a "traditional" classroom.

The ability to provide teachers with findings based on real world data gathered from hundreds of classrooms would provide instructors with the knowledge and confidence to flexibly adapt their learning process to a changing class composition. Only requiring a camera to implement, TailorEd could augment the experience of teachers in both common and more specialized classes by drawing statistical conclusions from hundreds of classrooms in different schools and providing recommendations based on real world results. TailorEd would allow governments to better assist commonly overlooked populations in education, allowing schools to tailor their education style to the student body, rather than attempting to copy what works for schools in different populations and expecting it to translate with no regard for differing needs and expectations.

2.2. Why use neural networks

The selection of AlexNet - a well established CNN for image recognition - as the CNN to categorize the classroom images helped alleviate initial concerns that the non-uniform angles of the photos, as well as significant variations in lighting and focus of the photos from classroom to classroom, would damage the accuracy. AlexNet is capable of identifying off-center objects [**?**] and also augments the data it receives by altering the intensities of RGB channels in the images it trains off of, in order to ensure that object identity is unchanged by different lighting [**?**].

A possible approach to determining the layout of a classroom is to determine the points on the image which correspond to the centers of each desk. A probabilistic shape-matching algorithm [6] can then determine which set of class shapes these point sets correspond to. However, AlexNet would only be able to detect round classroom configurations and not the other three since they depend on whether students' desks are facing the front of the classroom or each other and whether students even exist in the image.

For analyzing the direction faced by the students CCID is inspired by a group of students at Simon Fraser University who used recurrent neural networks to analyze relations in group activities happening in images [7] and created a structure inference machine capable of iteratively reasoning about which people are interacting in a given image as well as who is involved in a group activity. For instance, given a scene captured of a sidewalk, their machine is able to report which people are walking and which are waiting. This is accomplished by building a model that connects the low-level classifications to higher-level compositions [7]. In classroom image recognition, this method helps determine every individual student's action as well as their interaction with each other.

2.3. Activity analysis

The CCID project was designed to determine the change in classroom layouts of specialized classrooms with movable desks in order to determine the effect of these dynamic classrooms. The machine learning model was constructed using transfer learning on AlexNet, an early and lightweight CNN using a computer vision model (Krizhevsky et al., 2012).

However, while sufficient for differentiating between classroom layouts, determining what pedagogical activity was taking place in a class required a more specialized model than AlexNet. Developed in 2015 as the winning submission to the ImageNet Challenge, ResNet was the first model to surpass human accuracy in object recognition tests [2]. ResNet achieved these results by deepening the model to over 160 layers and combining that deepness with residual skip connections in order to resolve the problem of vanishing gradients in very deep models.

3. Methodology

Transfer learning is a machine learning method where a model developed for a task is used as the starting point for a model of a second task. As such, it is an optimization that allows rapid progress or improved performance when modeling a new task.

3.1. Classroom Configuration Identifier (CCID)

Given the vast computing and time resources required to develop neural network models, transfer learning is a practical solution for identifying the configuration of classrooms [4]. The specific Transfer Learning CNN chosen for use in TailorEd's Classroom Configuration Identifier (CCID) for detecting desks and people in classroom images was AlexNet. Which is an appropriate choice because it has been successfully used for object recognition [8] and its MATLAB implementation has been "trained on over a million images and can classify images into 1000 object categories" [9].

To ensure that trend results are traceable to specific choices made in the training of CCID, it is important to identify which parameters were changed and which were not. The accuracy of the network as the percentage of images identified correctly out of the total number of images was tested for and the confusion (error) matrices of key networks were extracted to be used for analysis.

Training Process:.

1. The size of the training set needed to be determined. Even though some classifications such as lecture had many more tagged images to train from, letting the network be trained with all of the sorted images would mean giving it a bias towards lecture classifications. Therefore, after finding the maximum amount of images which had been manually labeled by the team for each category, a constant number of 240 images per classroom layout classification was chosen to train the network.

2. The proportions of training-to-validation-to-testing sets were adjusted in order to find the optimal sizes which minimizes overfitting the training data set.

3. The batch sizes were varied to strike a balance as larger batch sizes allow the network to find key features amongst more images, while smaller batch sizes allow for faster processing.

4. A balance was struck for the learning rate, as a low learning rate ensures a network will not miss any small local minima, while a high learning rate allows for faster training.

5. And finally, since choosing the best algorithm for reaching optimal weights is pivotal to this network's success, MATLAB's most popular options of Stochastic Gradient Descent with Momentum (SGDM), Adaptive Moment estimation (ADAM), and Root Mean Square Propogation (RMSProp) were interchanged in order to find the most optimal optimization algorithm for CCID's Alexnet. ADAM and RMSProp are derived from SGDM and change the learning rate of the network based on how close it is getting to a solution[10].

3.2. Classroom Activity Identifier (CAID)

Building a Classroom Activity Identifier (CAID) was broken down into two phases:

(i) The human coding phase, where trained human coders label the activity going on in each image manually

(ii) The machine learning phase where a neural network trained, tested, and validated using these images to automate the analysis of aerial classroom photos according to captured activity

Qualitative coding of classroom photos for activity. In order to ensure that the images were coded accurately and similarly by all human coders, the coders began with a basic list of activities and coded 100 photos over three months as part of a coder norming process: Every week,

- each coder categorized a shared set of 100 images, randomly selected from the photo corpus.

- then coders met to discuss categories and category definitions, focusing on images that had been labeled differently by different coders.

At the conclusion of the photo labeling training, the human coders achieved an overall agreement rate of 80% around the following classroom activity categories:

- Empty: no/too-few students are present in classroom

- Lecture: students are receiving information from a single presenter (including instructor lecture, student presentation, and/or viewing media)

- Discussion: students are engaged in common discussion activity, attending to a member of the class (not a presenter situated away from the student group)

- Group work: students are formed into small groups working on a shared task (including discussing, reading, and/or writing)

- Writing (solo): students are writing individually, either by hand or typing on a device

- Reading (solo): students are reading individually, either from print materials or digital devices

Originally, several more categories of classroom activities were included such as a break down of group work activity or the determination of whether students were speaking in pairs. However, examples of these categories were too rare in the captured dataset. As a result, they were condensed into the general group work and discussion categories respectively.

For six months following this norming process, each week coders received 50 images randomly selected from across the entire photo corpus to label. Each image was assigned to two randomly paired coders, with pairings changing from week to week. If both coders labeled the image the same way, its categorization was final. If the coders disagreed, the image was submitted to a third coder for a "tie-breaker" vote. If the third coder picked yet a different label for the image, TailorEd's Principle Investigators (PIs) used the category definitions to assign a final label to the image. These "problematic" images were tagged as ambiguous and not used to train the neural network, though kept in the corpus for future analysis. The 3,700 non-ambiguous images labeled by the human coders were then used to train CAID, which was then used to label 18,000 photos analyzed by the time of this writing.

Neural network categorization of classroom photos. The AI phase of this project involved two stages: standardizing images to account for inconsistencies in photography and use of the CCID and CAID transfer learning CCN classifiers.

To standardize the images taken in the nine different classrooms, MATLAB image processing capabilities were used to regularize the classroom images, specifically to deal with differences in photos caused by differences in camera lenses, lighting, and focus (visible in the raw photos in Figure 1). After standardizing the photos, transfer learning the trained CCID and CAID models were applied to the images in order to label all images.

Furthermore, Hyperparameter tuning was applied to find the optimal learning rate, batch size, and number of epochs to maximize the model's accuracy. All permutations were examined for the initial learning rates of 0.001, 0.0025, 0.005, and 0.01 alongside batch sizes of 10, 20, 40, 50, and 100 images. The best result proved to be a combination of an initial learning rate of 0.001 and a batch size of 20. Furthermore, the model was trained over thirty epochs in order to achieve the highest level of accuracy of the model.

Due to imbalance in the dataset, a Self-supervised training method was adapted to take advantage of unlabeled data. ResNet was set up to train under the MoCo v2 setting [1] as will be discussed more in section 4.2

4. Results

4.1. CCID

In the final network, 70% of the data was selected to be training data, 20% was used for validation, and 10% was used for testing. This means that out of the 240 images for each classification, 168 images were in the training set, 48 were chosen for the validation set, and 24 were used for testing. In total, out of the 960 images used for training, 672 images were in the training set, 192 were in the validation set, and 96 were in the testing set.

After trying all combinations in which the mini-batch size was set to 5, 10, 15, and 20 and the initial learning rate was set to 0.01, 0.005, and 0.001, the highest level of accuracy was achieve with a mini-batch size of 20 and a learning rate of 0.001. This is due to the mini-batch size of 20 allowing for more images to be trained upon in every iteration of training where the initial learning rate of 0.001 ensures that the weights are not changed too haphazardly.

Using these settings for the CNN, an accuracy of 97% was thus obtained for classifying 1364 images.

Figure 2 shows the confusion matrix of CCID. Each cell depicts how many entries of the true class were accurately predicted. The numbers along the blue diagonal are the number of images correctly identified by CCID. The two darker orange cells highlight the aforementioned discrepancy of when CCID misidentifies a classroom as a group formation. For instance, 6 empty classrooms were misidentified as groups and 13 lectures were misidentified as groups.

		Numbers			
		Predicted Class			
		Empty	Groups	Lecture	Round
True Class	Empty	231	6	5	0
	Groups	1	268	4	2
	Lecture	1	13	556	1
	Round	0	2	1	273

Figure 2. Confusion matrix of CCID (Number of images identified)

Figure 3 shows the same confusion matrix as Figure 2, but with the percentage of the true class identified as the predicted class in each cell. The table shows that the round class was the easiest to correctly identify (with 98.91% of the true class identified correctly) which is due to the unique shape of a round classroom. Specifically, a round classroom comes in either the shape of a circle or a U while the empty, lecture, and group formations could come in many different shapes. Also, empty classes were the most difficult to correctly identify with 95.45% of the true class identified correctly.

		Percentages of true class			
		Predicted Class			
		Empty	Groups	Lecture	Round
True Class	Empty	95.45%	2.48%	2.07%	0.00%
	Groups	0.36%	97.45%	1.45%	0.73%
	Lecture	0.18%	2.28%	97.37%	0.18%
	Round	0.00%	0.72%	0.36%	98.91%

Figure 3. Confusion matrix of CCID (Percentages of true class identified)

Figure 4 shows the same confusion matrix as Figures 2 and 3, but with the percentage of the predicted class identified to be in that predicted class in each cell. This table allows us to judge the confidence of the network in its classification of an image. If the percentage of a correctly identified predicted class is high, then when CCID identifies an image to be of that class, it is more likely that the identification is correct. For example, 99.14% of the images that were predicted to be of an empty classroom were correctly identified as empty. However, only 92.73% of the images classified as groups were actually groups and 2.08% and 4.50% of the images classified as groups were actually empty

and lecture formations respectively with a final 0.69% classified as round. This is because some of the images that are identified to show a class doing group work portray a few students who had only turned their heads to the side rather than having rotated their entire desks, which leads to features that are much harder to detect. Therefore, when CCID classifies an image as empty, it has a higher likelihood of being correct than when it classifies an image as a group.

An interesting result is that if a classroom is classified as empty, this classification has as high as a 99% chance of being accurate, which is the highest chance of accuracy that the network could achieve with any particular classification. Overall, however, only 95.45% of the empty classrooms were classified correctly. This percentage is lower than that of the other three groups by around 2%. Thus the empty room classification is the hardest one to be predicted even though the network marks this classification with the highest confidence. One of the reasons that the network misses many of the empty classrooms is the fact that the empty classrooms can be arranged in any shape with the desks facing in any direction. At the same time, the amount of confidence that the network has with the empty classroom classification is understandable and reasonable because no people are in these images. Therefore, the color variation and contrast in smaller sections of the images will be significantly less. Figure 5

Percentages of predicted class

		Predicted Class			
		Empty	Groups	Lecture	Round
True Class	Empty	99.14%	2.08%	0.88%	0.00%
	Groups	0.43%	92.73%	0.71%	0.72%
	Lecture	0.43%	4.50%	98.23%	0.36%
	Round	0.00%	0.69%	0.18%	98.91%

Figure 4. Confusion matrix of CCID (Percentages of predicted class identified)

shows the confusion matrix of two other key networks that failed.

One network used ADAM while the other used a higher learning rate of 0.01. ADAM is an adaptive method of calculating changes. It changes the learning rate of the network as it goes through the examples. With a fluctuating and/or higher learning rate, the network's weights get changed much more aggressively, which can make it easy to skip local minima while trying to optimize the weights of the network. This means that there is a large chance that the network will miss more subtle features in the images. Such subtleties are key for detecting the difference between some classifications, such as the direction students face in lecture formations versus the direction they face in group formations. Both networks result in the same confusion matrix that predicts all images to be lecture, reaching a similar accuracy around 41%.

Numbers

		Predicted Class			
		Empty	Groups	Lecture	Round
True Class	Empty	0	0	242	0
	Groups	0	0	275	0
	Lecture	0	0	571	0
	Round	0	0	276	0

Figure 5. Confusion matrix of failing key networks (Number of images identified)

4.2. CAID

Using CAID, several different combinations of hyper-parameters were examined with the goal of selecting not only the highest overall accuracy but also the best distributed accuracy. This distinction is made due to the fact that a balanced sample of all image categories was not able to be obtained for training and testing. Several categories, such as discussion or reading, had very few examples with which to train and test the model off of. Many classes exclusively consist of only lecturing and empty classroom images. As a result of this unbalanced dataset, an unbalanced training and testing set had to be accepted as otherwise the model would be retrained on a small dataset. This imbalance also extended to the testing set, meaning that if the model performed well on the most common classes, then poor performance in identifying other less common activities could be hidden in the overall accuracy. In order to alleviate the problem, the dataset was augmented by selecting images in the smaller categories and randomly rotating or translating them across the x and y axis in order to generate more examples to test the model off of. However, even with these augmentations, the size of the training set remained small.

Early experiments with hyper-parameter tuning attempted a large batch size of 100 for the model. The theory behind the large batch size was that if memory was not an issue, then more images should be provided to the model on every batch in order to better learn distinctions between them. However, as can be seen in Figure 6, the model still had the issue of performing with a high level of accuracy on categories with many examples but poorly on categories without. The accuracy for the high batch size model was 79.82% and was improved upon by future models. Another attempt that improved can be seen in Figure 7 where a batch size of 20 and a large initial learning rate of 0.01 were selected. While the model did perform well with an accuracy of 82.53%, it still maintained the same

strengths and weaknesses as before. Almost all of the improvement made in the model can be seen in the *Writing* category, where it became capable of accurately predicting most Writing classroom photos.

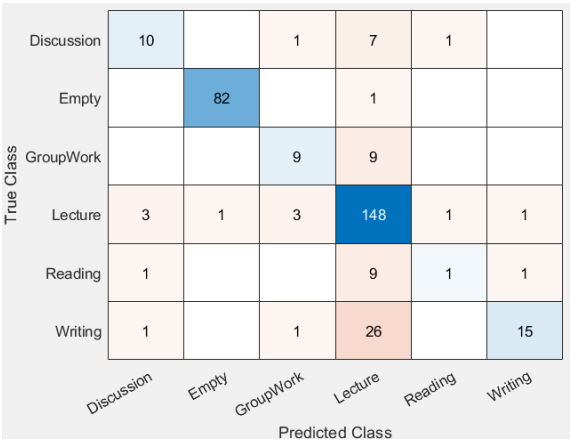

Figure 6. 100 Batch Size Confusion Matrix

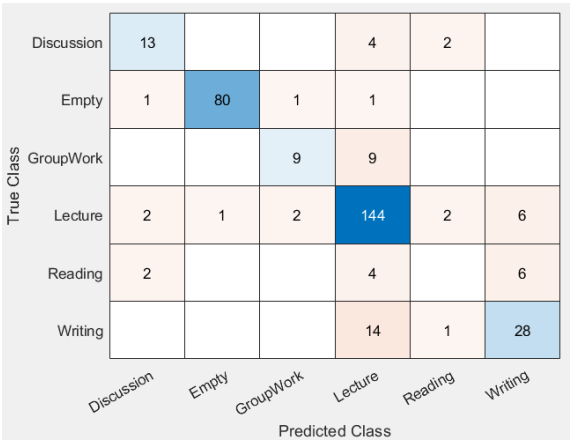

Figure 7. 0.01 Learning Rate Confusion Matrix

The ultimate result was a model that could accurately predict the pedagogical activity occurring in the class 86.17% of the time. It should be noted, however, that this accuracy result doesn't tell the whole story, illustrated in Figure 8. Because the model was trained on a representative sample of pictures from classrooms, the number of images depicting each activity varies considerably because some activities are used much more frequently than others. The number of images depicting each activity affects the accuracy rate of each category, with activities with fewer examples having lower accuracy rates because CAID can't generalize the category parameters as effectively. For example, because by far the greatest number of images show lecture activities, this category has a very high accuracy rate. For activities that are rarely depicted in photos (like reading), the accuracy is lower.

Figure 8. Final ResNet Class Activity Confusion Matrix

One highly promising trend however is that in classes with very high numbers of examples, such as *Empty* and *Lecture*, the model's performance is nearly perfect. In the case of predicting empty classrooms, the model accurately identified every single empty classroom correctly and only classified one other classroom as being empty erroneously. While empty and populated classrooms may be easy to distinguish between, the trend also showed the same for lecture vs. non-lecture activities. This trend suggests that if the CAID model was provided with a properly sized and equally balanced dataset, then the accuracy of the model would be competitive with the CCID model.

One final attempt to improve accuracy with the limited dataset was attempted using ResNet trained under a MoCov2 setting as a self supervised model. Different data augmentation techniques were explored with this option, including random-crop, color jitters and Gaussian blur in order to attempt to create a larger augmented dataset. Ultimately however, the result of the prediction was unable to exceed 84% accuracy, under-performing the original ResNet CAID model. One of the reasons for this could perhaps be that the augmentation techniques are inferior to the original rotation and translation augmentations. The reason for the difference in augmentation performance could be that the original ones produce images that more closely match the training set images as it simply shifts pixel locations rather than changing their values with respect to nearby pixels as would occur in blurring or color jitters. Furthermore, even though self-supervised learning can leverage unlabeled data, it still likely requires more initial data in order to be enough to fine-tune the network to have a better representation for classification.

Ultimately, it can be determined that the limited dataset and unbalanced focus on lecturing in the dataset is a consequence of data gathering from a single university. In order for CAID and the TailorEd project as a whole to accomplish its stated goals, data gathering will need to be expanded beyond Santa Clara University. Universities each develop their own particular culture surrounding instruction, and while the results of these models may generalize well for Santa Clara University, different institutions will have differing emphasis on styles of teaching. By gathering data from other Universities, the models will become less specialized and more robust, providing more accurate predictions that can be used to better understand how classroom environments and classroom activities impact the educational outcomes of their students.

5. Future Work

5.1. Creation of Class Activity Maps

In an effort to create a visualization of the usage of time spent on each activity type in a classroom, a color-coded activity map is under development which will utilize the output from the CAID classifier as an input. Many designs are currently under user testing in order to determine which will be most useful and easiest to decipher information from.

5.2. Creation of Classroom Utilization Heat Maps

In an effort to create visualizations of the usage of space within dynamic learning classrooms, the generation of a heat map for each classroom is underway. In addition to tracking the regular movement of students and desks across the space of the classroom, these heat maps can be coded with several different layers. These layers can provide information on what activity the spaces are commonly used for, providing teachers and especially classroom designers with a better understanding of how the layout of a classroom is being utilized or perhaps even underutilized.

5.3. Gathering Additional Higher Quality Images

One of the most immediate future steps will be to continue gathering more images, particularly on the least common classroom activities. Unfortunately, due to the COVID-19 pandemic only recently have classes at Santa Clara University begun to meet in person again, thus halting image collection until Fall quarter of 2021. However, the period without classes did allow for the construction of new facilities to progress faster. As part of the re-imagination and construction of the school of Engineering, all of the old facilities were demolished and a new building was constructed which opened for in person instruction just as in person teaching was allowed again by federal and state rules for the fall quarter of 2021. Due to the demolition of old facilities, several classrooms in the study were lost. However, due to generous funding from the Academic Technologies of the University's Information Services department, all 9 newly constructed classrooms were added to the study via installation of new cameras, bringing the total count of classrooms within the study to 27, more than doubling data collection capacity. This opportunity also allowed for the installation of the new cameras more centrally in each classroom, which allows for capturing images of the front of the classroom which earlier camera placements often obscured. This enables the determination of the position of the instructor and the use (or not) blackboards/whiteboards and projectors, which help better deduce the activity taking place in the classroom during the capturing of each image.

5.4. Creation of a Classroom Technology Identifier (CTID)

Due to the aforementioned relocation of cameras to a more central location in the classrooms, an additional opportunity for the study of classroom technology presented itself. The images can be used to train a new model which can identify the various technologies such as blackboards/whiteboards, projectors, laptops, and mobile phones in use during a class. This will enable future consideration of the relationship between these technologies and student learning outcomes.

Unlike previous applications of the TailorEd classifiers, however, the technology classification categories will not be mutually exclusive as multiple different types of technologies can be at use in a classroom at the same time.

6. Conclusion

This paper explores the extent to which the problem of pedagogical research being reliant on small sample sizes or inaccurate self-reported data can be addressed using big data and machine learning. CNNs continue to be the best performing ML models for use with computer vision applications, and can be easily adapted to new tasks through the use of transfer learning. But unfortunately, transfer learning requires large and balanced datasets in order to create a strongly performing model. For the CCID model, a large and balanced dataset is present, resulting in a highly accurate classifier able to distinguish between four different classroom configurations with over 95% accuracy. However, in the case of CAID, the dataset is not yet adequate for determining all classroom activity types.

TailorEd attempts to resolve the issue of having only small amounts of data for some activity categories

by using a self-supervised model as well as data augmentation to enhance the dataset. While ultimately the self-supervised model proved unable to advance the results past the accuracy of the data augmented model, improvements were made upon the base dataset. The resulting CAID model has shown excellent performance in classes with large numbers of examples, but reduced performance in classes with limited numbers of examples. Still, the 86.17% overall accuracy obtained by CAID shows that the model can be useful for classifying large numbers of classroom images. Given the new flood of higher quality images being gathered, however, CAID's accuracy will undoubtedly rise.

Acknowledgement. Many thanks are due to Nancy Cutler, Deputy Chief Information Officer (CIO) for Academic Technology and the Co-Director of the Collaborative for Faculty Innovation, for obtaining necessary funding for the project, and to Joel Bennett and the Media Services Group of the Academic Technology unit of Santa Clara University's Information Services for continued technical support of the project in the form of camera installation and image warehousing. Thanks also for the continued support for this project from the Department of Mathematics & Computer Science (MCS) and the Department English of the SCU College of Arts & Sciences and from the Department of Computer Science and Engineering (CSE) and the Frugal Innovation Hub (FIH) of the SCU School of Engineering. Finally, we thank other past and present TailorEd team members for their contributions: Stefan D'Costa, Meghan McGinnis, Oras Phongpanangam, Mohammed Khadadeh, Liying Liang, Amanda Taylor, Mika Philip, Jacqueline Reardon, Maximilian Khan, Roland Afaga, Jay Ladhad, Laurynn Diby, Ashika Rajesh, Carson Hom, Vivek Ponnala, Axel Perales, Jiahong Li, Juhi Checker, Shiv Jhalani, and Luis Herrera. Finally, this work would not have been possible without the CAID photo labeling student research assistants: Liam Abbate, Roland Afaga, Laurynn Diby, Katya Keklikian, Kristin Lee, Justin Ling, Ann Nguyen, Julia Perry, Kishann Rai, Ashika Rajesh, Gabe Reed, Malika Singh, Fiona Sundy, Sneha Vinod, Morgan Yazdi, and Kristina Yin.

References

[1] Chen, X., Fan, H., Girshick, R., & He, K. (2020). Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference On, , 770-778. https://doi.org/10.1109/CVPR.2016.90

[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, 60(6), 84-90. https://doi.org/10.1145/3065386

[4] Shaghaghi, N., Khadadeh, M., McGinnis, M., Liang, L., & Calle, A. (2019). Classroom Configuration Identifier (CCID). 2019 IEEE 11th International Conference on Engineering Education (ICEED), , 204-209. https://doi.org/10.1109/ICEED47294.2019.8994827

[5] R. Wannarka and K. Ruhl, "Seating arrangements that promote positive academic and behavioural outcomes:A review of empirical research, "Support for learning, vol. 23, no. 2, pp. 89–93, 2008.

[6] G. McNeill and S. Vijayakumar, "A probabilistic approach to robust shape matching," in 2006 International Conference on Image Processing. IEEE, 2006, pp. 937–940.

[7] Z. Deng, A. Vahdat, H. Hu, and G. Mori, "Structure inference machines: Recurrent neural networks for analyzing relations in group activity recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4772–4781.

[8] A. Uar, Y. Demir, and C. Gzeli, "Object recognition and detection with deep learning for autonomous driving applications,"Simulation, vol. 93, no. 9, pp. 759–769, 2017.

[9] MathWorks. Transfer learning using alexnet.

[10] Y.-H. Byeon, S.-B. Pan, and K.-C. Kwak, "Intelligent deep models based on scalograms of electrocardiogram signals for biometrics,"Sensors, vol. 19, p. 935, 2019.