Scalable Mind-Wandering Detection for MOOCs: A Webcam-Based Approach

Yue Zhao, Christoph Lofi, and Claudia Hauff

Delft University of Technology, Web Information Systems, Delft, the Netherlands
{y.zhao-1,c.lofi,c.hauff}@tudelft.nl

Abstract. Mind-wandering and loss of focus is a frequently occurring experience for many learners, negatively impacting the learning outcomes for activities like reading materials or attending lectures. While in a classroom setting, a skilled teacher might be able to detect mind-wandering in her students and to intervene, which is hard for online classes and Massive Open Online Courses (MOOCs). However, previous studies suggest a strong relationship between learners' mind-wandering and their gaze movement, allowing to detect mind-wandering in real-time using eye tracker devices. Unfortunately, current research in this area relies on using specialized professional hardware, and thus cannot be used in real-life online learning or MOOC scenarios due to the inability to scale beyond lab settings. Therefore, in this paper, we propose a method using ubiquitously available consumer grade webcams for detecting mind-wandering in learners. In a controlled lab study, we compare the results of webcam-based approaches with those using professional eye trackers, and based on our results argue that a large-scale application in MOOCs is indeed possible.

Keywords: learning analytics, MOOCs, learner attention, mind-wandering, eye gaze tracking

1 Introduction

Mind-wandering is an essential part of human behavior consuming up to 50% of everyday thoughts [6], and can be described as "thoughts and images that arise when attention drifts away from external tasks and perceptual input toward a more private, internal stream of consciousness" [9]. While generally mind-wandering can also have positive effects like fostering creativity [16], many educational tasks like following a lecture or reading teaching material require active attention and focus to reach the desired learning outcomes. For these tasks, excessive mind-wandering can have disastrous effects on learning efficiency [15].

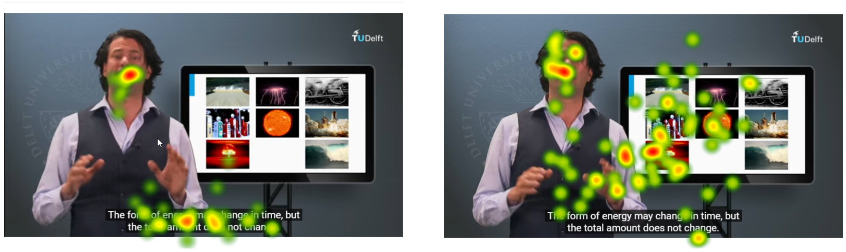
Mind-wandering and attention lapse in traditional classroom settings have been studied for a long time. In [8], McKeachie et al propose that learners may lose their attention during lectures quite quickly. For example, Bunce et al. [3] ask students to self-report mind-wandering during lectures by pressing a button on their desk, and find that students' attention usually cycles several times within a 9-12 min course segment. However, the actual attention span of learners is still contested, and in [17], Wilson and Korn argue that the often cited 10-15 min attention span is not yet well enough supported by previous literature.

For online courses and MOOCs, this problem is even more severe as they are consumed using digital display devices. Such consumption mode is particularly prone to mind-wandering. Likely due to the ubiquity of smart phones and digital content, a significant subgroup of online users adopt a "heavy media multitasking" behavior [7], making it especially challenging for them to exclusively focus a single multimedia content unit. This is also supported by our studies in this paper, where learners frequently lose focus even in short video clips of around 7 minutes.

For large-scale research beyond lab studies into the issue of mind-wandering when consuming digital educational materials, an automated and unobtrusive method is necessary which does not actively interrupt and distract learners (by e.g. asking them to press a button). In addition to providing insights into learner's behavior, such methods would also allow for real-time interventions to boost learning effectiveness. For example, should a MOOC video player detect that a learner loses focus and starts mind-wandering, the video could be stopped in order to avoid skipping over relevant content, or a visual or acoustic signal could be used to bring a learner's attention back.

To this end, gaze tracking seems to be a promising contender, as previous research shows that by analyzing eye movement and gaze measures, mind-wandering can be detected. This can be attributed to the eye-mind link effect [12], stating that "there is no appreciable lag between what is fixated and what is processed". Previous studies successfully detect mind-wandering using gaze data for reading textual material on screen [1], and for watching (non-educational) films [2]. However, these and similar works usually rely on using expensive and specialized eye tracking hardware, which is typically not available to the average MOOC learner. Thus, neither large scale studies nor mind-wandering-based interventions in real-life MOOCs can be performed using these methods.

Therefore, in this paper, our goal is to develop a fully automatic method for detecting mind-wandering and loss of focus in near-realtime using only low-end webcams ubiquitously found on laptop computers or used for computer-based video chatting. To motivate this approach, refer to Figure 1, showing an excerpt from the study conducted in this paper. The figure depicts the heatmaps of the gaze fixations of two participants in 30 seconds time interval. The shown MOOC video has several relevant visual areas, as for example the lecture slides, the subtitles, or the speaker's face. In the depicted scene, a changing set of examples are shown on the slides which would be relevant to grasp the lecture's content. The participant who reports mind-wandering intently gazes on a spot on the speakers face, ignoring the slides and the shown examples, while the second participant who reports no mind-wandering actively focuses on all relevant areas of the video. Our method relies on automatically learning such mind-wandering



(a) reported mind-wandering (b) no mind-wandering

Fig. 1: Gaze heatmaps of two participants over a 30s interval.

patterns based on the combination of simple video features and the recorded gaze features obtained in a controlled laboratory study from 13 participants, using self-reported mind-wandering as a reference baseline for mind-wandering detection. In summary, our goals are as follows:

– We create and provide an elaborately generated gold dataset for fostering future gaze-tracking-based mind-wandering research, featuring 13 participants watching two MOOC videos each in a controlled laboratory setting, reporting feedback on mind-wandering in brief intervals. In addition to these mind-wandering reports, we provide video and gaze data as recorded by a specialised professional eye tracker as well as gaze data recorded by webcam-based solution. This data is openly available on our companion web page [18].
– We implement and evaluate a method for automatically detecting mind-wandering using professional eye tackers, relying on the results and best practices published in [2].
– We design a method for automatically detecting mind-wandering relying only on a simple webcam and open-source gaze analytics.
– We extensively discuss and evaluate both approaches, and argue that our webcam-based method is indeed suitable for large-scale research outside a controlled laboratory setting.

2 Methodology

In this section, we describe the overall design of our study, first focusing on how we design our experiment to collect the ground truth of learners' mind-wandering and their gaze estimation during watching lecture videos. After the introduction of the experiments, we discuss how to select features from eye gaze data and how to detect learners' mind-wandering based on these features.

2.1 Experiment Setup

In our experiments, we focus on gaze data, i.e. the points and areas on the screen a participant is actively looking at. Gaze data cannot be measured directly, but is estimated from eye and iris movements using different techniques and algorithms by the software or hardware frameworks used in this study. As the first step, we obtained such gaze data from 13 participant in controlled laboratory-based experiment. During this experiment, we also obtain the mind-wandering ground truth by relying on periodic self-reporting of the participants.

Our experiment includes two introductory videos taken from two MOOCs offered by TU Delft on the edX platform. One video covers the basics of the atomic model taken from the MOOC "Understanding Nuclear Energy", while the other video is on energy conversion from the MOOC "Solar Energy". Both videos were selected because they use visually rich lecture slides overlayed with the speaker (see Figure 1). Furthermore, they are on topics which we consider interesting to a wider audience and do not require extensive pre-knowledge to understand their content. Both videos are x-MOOCs [13], and as such are carefully and professionally produced, and are quite short (6m41s and 7m49s respectively) to keep learners focused and interested. For controlling any effects which might result from the order in which these videos are presented to our participants, one randomly selected half of the participants will first be shown the nuclear energy video followed by the solar video, while the other half of the participants sees a reversed order.

For gaze estimation, as a reference, we used the professional Tobii X2-30 eye tracker and corresponding software Tobii Studio, a specialized device which can obtain high-quality eye tracking data with very high sampling rates, thus resulting in very precise gaze estimation data. For our webcam-based approach, we used the camera integrated in the machine used during the experiment, a Dell Inspiron 5759 laptop computer with a 17-inch screen and a 1920×1080 resolution. For estimating gaze data from a live webcam feed, we used the WebGazer.js [11] open source framework. Participants were asked to sit stable and comfortable in front of the screen, and the distance between eyes and screen was between 52 to 68 cm. We used normal office light conditions in our experiment, consisting of natural daylight and normal office ceiling light.

In our laboratory study, we have 13 participants in total (6 females and 7 males) who volunteered from a pool of computer science MSc and PhD students. 6 participants are wearing glasses or contact lenses, which is often challenging for eye tracking. Their gaze is estimated simultaneously by both WebGazer and the Tobii eye tracker while watching the MOOC videos. The data generated by Tobii studio includes the estimated 2D gaze coordinates for each eye, the duration and coordinates of gaze events (i.e. fixations and saccades), eye position of participants, distances between participants and camera, pupil position etc with a sample rate of 20 samples/sec. However, the data extracted from our webcam-based eye tracking only includes the estimated 2D gaze coordinates of both eyes combined with a sample rate of 5 samples/sec. We use a custom-built web application closely resembling real MOOC lecture video players. This application also records all gaze data. Each participant is orally instructed by the experiment's supervisor, and the web application contains a brief tutorial

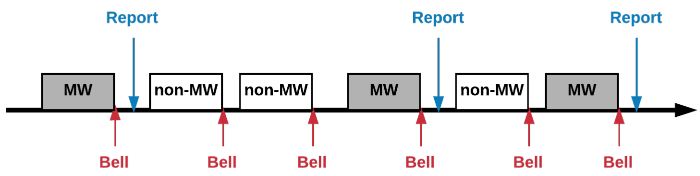


Fig. 2: An example mind-wandering report

which familiarizes each participant with the controls of the application before starting the experiments.

In order to both train and benchmark our approach, reliable reference data is required with respect to when the participants actually focus on the shown MOOC videos, and when they are mind-wandering. For this, we adopted the experimental design relying on periodic self-reporting which was already successfully used in previous studies, like for example [1, 7]. This design relies on using a so-called "bell" signal (which, in our case, is a pleasing medium-volume acoustic bell signal played by the web application). After the bell, participants may report if they have been mind-wandering during the time since the last bell signal (or not) by pressing a feedback button in the web application. Typically, the next bell signal will be given after another 30-60 seconds. The actual time is randomized within those bounds, as previous research [1,7] suggests that participants perceive interruptions which are not perfectly periodic as less interrupting. In order to further limit the mental annoyance of this process, participants were only asked to actively report in case they have indeed been mind-wandering. If they stayed focused during last bell time period, the bell signal may be ignored and no report is necessary. This process results in mind-wandering reports for each participants, including the bell signals and participant responses with respect to mind-wandering as shown in Figure 2.

2.2 Mind-wandering Detection using Eye Gaze Features

To realize mind-wandering detection using gaze tracking using the professional eye tracker and our webcam-based solution, we rely on supervised machine learning classifiers which are trained using the aforementioned mind-wandering reports as reference labels, and aggregated gaze features for the time spans between two bell signals as collected by either technique for classification input.

For the Tobii gaze data, inspired by [1,2] we cover 58 features in total. These features can be classified into two groups, global features and local features. The global features refer to features which are independent of the current content of the MOOC video, and are as shown in Table 1 based on fixations (i.e., maintaining of the visual gaze on a single location) and saccades (i.e., the quick and simultaneous movement of both eyes between two or more phases of fixation in

Table 1: Features leveraged in the detection of participants' mind-wandering

Feature Name	**Explanation**
	Global Features
Fixation Duration	the durations (ms) of fixations
Saccade Duration	the durations (ms) of saccades
Saccade Distance	the distances (pixel) of saccades
Saccade Angle	the angles (degree) between saccades and the horizon
Number of Saccade	total number of saccades
Horizontal Saccade Ratio	the proportion of the number of saccades which have saccade angles less than 30 degree
Fixation Saccade Ration	the ratio of the durations of fixations to the duration of saccades
	Local Features
Saccade Landing	the proportion of the number of saccades landing in different areas
Fixation Duration AOI	the durations (ms) of fixations located in different areas

the same direction). The feature vector of a given bell time span covers statistical aggregates of fixation and saccade data, like maximum, minimum, mean, median, standard deviation, range, kurtosis and skew of fixation durations, saccade durations, saccade distance and saccade angles.

Local features are mainly based on the relationship between fixations/saccades and the areas of interest (AOIs) in the MOOC video, i.e. local features correlate gaze data with the current video content (e.g., similar to our reasoning in the introduction, there are certain areas of a video where a focused learner should focus her attention in order to follow the content, while others are less interesting). While this opens a large and complex design space for engineering features, we opted for a very simplistic implementation in which we manually define three fixed areas of interest, namely the instructor's face, the video subtitles, and the content area showing the lecture slides. Resulting local features are for example then the number and/or length of saccades and fixations which focus on different areas for a given bell time span. All saccade and fixation data is aggregated by Tobii studio with high precision for each bell time span based on a raw sample rate of 20HZ.

In contrast, due to technical limitations of the used WebGazer gaze tracking framework, we can only use a sample rate of 5HZ for our webcam-based experiments. As changes of eye fixations and saccades usually happen within the range of 200ms to 400ms [14], reliable gaze data comparable to the one provided by the high-speed Tobii tracker is hard to obtain using such a low sample rate and thus needs to be estimated algorithmically. For this, we rely on micro-saccade detection as discussed in [5]. The saccades between two consecutive gaze estimations are calculated based on the velocity of eye movements. The gaze estimations between two saccades are treated as one fixation. If there is only one gaze estimation between two saccades, we assume this gaze estimation is a fixation with a duration between this gaze estimation and the next gaze estimation. After the detection of saccades and fixation, we can generate the same 58 features as already shown Table 1 also for webcam-based eye tracking. Likely, the resulting feature vectors are less precise, but we will show later that they still show comparable classification performance as we rely on aggregating features over the time spans between the bells, thus the smaller imprecisions carries little weight.

For training our supervised machine learning model, we adopt a leave-one-participant-out cross validation method as discussed in [10]. In each run, the data of one participant is selected as the test data and the data of all other participants is used as training data. Based on previous work [1, 2, 10], the data on learners' mind-wandering is usually unbalanced. The mind-wandering rate usually range from 0.2 to 0.35. We counter the effects of the imbalance by using the oversampling method Synthetic Minority Over-sampling Technique (SMOTE) as described in [4]. We tested different supervised classifiers like L1/L2-Logistic Regression, Decision Trees, and Gaussian Naive Bayes.

All our collected data including the used lecture videos, the participant's mind-wandering reports, and the raw and processed data and features obtained from the Tobii eye tracker and the WebGazer webcam system are available on this paper's companion webpage [18] to stimulate and foster further research on this topic.

In order to get the best results for the mind-wandering detection, we use different subsets of all features in our experiments, which are global features only (G), local features only (L) and the combination of global and local features (G+L). Considering also the use of SMOTE on unbalanced data, we can train and evaluate our model with 6 different setups of features and training-data pre-processing.

3 Results

In this section, we focus on the experimental results of our aforementioned study and described mind-wandering detection methods. We address three main research questions:

1. How do the participants mind-wandering reports look like, and what can be learned from them?
2. How well does our eye-tracker-based mind-wandering detection method perform?
3. How well does our webcam-based mind-wandering detection method perform, and how does it compare to using a professional eye tracker?

3.1 Exploratory Analysis of Mind-Wandering Reports

In this brief exploratory analysis, we analyse our participant's mind-wandering behaviour while watching the MOOC videos in our controlled lab experiment.

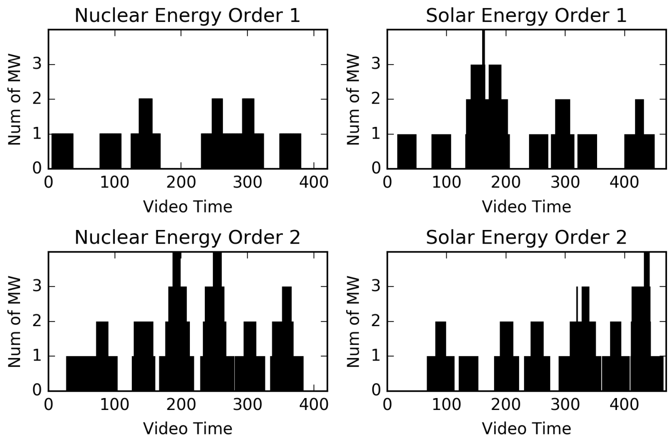


Fig. 3: The number of mind-wandering in each video with different order

There are only very few previous studies on learners' mind-wandering while watching lecture videos, as for example [7] which uses only 4 bell signals for two 18-min videos, and thus sheds little light on mind-wandering for short video-based lectures.

In Figure 3, the distributions of participants' reported mind-wandering events over the course of each of the two videos are shown. As briefly discussed in the last section, participants are shown both videos in a random order, which is also reflected in the diagram (i.e., Video Name Order 1 means that the respective video has been shown first, order 2 means it has been shown second.)

As the number of participants in each of the experimental groups is very small, no statistically significant conclusions can be drawn. However, it is visible that mind-wandering is indeed a rather frequent occurrence even for very short video lectures of roughly 7 minutes (our measured mind-wandering rate is 29%; i.e. in 71% of all bell time spans, our subjects actually stayed focused), and it seems that our participants tire considerably during the second video when the experiment draws to its conclusion. This feedback was pro-actively provided by several of our participants in a post-experiment questionnaire, and seems to be at least anecdotally confirmed by the presented mind-wandering report statistics.

3.2 Mind-Wandering Detection

In order to answer research questions 2 and 3, we investigate how accurately we can detect participants' mind-wandering based on gaze data extracted by Tobii and WebGazer. We show the results for all feature sets described in 2.2 in Table 2, including feature sets with/without the SMOTE method. For the sake of brevity, we only list the best result of 5 classifiers for each feature set. However, the full results as well as all parameters of the classifier algorithms

Table 2: Detection results based on gaze data. *G* means global features and *L* means local features.

Data	*Feature*	*SMOTE*	*Precision*	*Recall*	*F1*	*Classifier*
Baseline	–	–	0.290	0.291	0.290	–
Tobii Data	G	–	0.309	0.362	0.333	GaussianNB
	G	✓	**0.373**	**0.534**	**0.440**	L2-LR
	L	–	0.361	0.379	0.370	DecisionTree
	L	✓	0.299	0.500	0.374	GaussianNB
	G+L	–	0.311	0.397	0.348	DecisionTree
	G+L	✓	0.329	0.397	0.359	L2-LR
WebGazer Data	G	–	0.339	0.690	0.455	GaussianNB
	G	✓	0.345	0.707	0.463	GaussianNB
	L	–	0.342	0.690	0.457	GaussianNB
	L	✓	0.346	**0.776**	0.479	GaussianNB
	G+L	–	0.352	**0.776**	**0.484**	GaussianNB
	G+L	✓	**0.362**	0.724	0.483	GaussianNB

used can be obtained from our companion website [18]. As a *baseline method*, we used a random guessing classifier which includes the knowledge that the mind-wandering rate is 0.29 (i.e. presented with a feature vector, randomly in 29% all cases it will classify it as mind-wandering). Since accuracy is not a suitable metric for our unbalanced data, precision, recall and F1-measure are used in this experiment. *Precision* indicates the accuracy of our detection if we detect if a participant has mind-wandered in last 30 seconds, and *Recall* indicates how many actual mind-wandering events of a participant are detected.

Based on our results in Table 2, all our methods can generate results which are significantly better than *the baseline* on both *Precision*, *Recall*, and *F1 measures.* We can find that *Precision* in each methods with different data does not show any clear differences. However, applying the SMOTE method on Tobii data can increase *Recall* when we use global features or local features only. Both *Precision* and *Recall* do not have clear differences with respect to WebGazer features. All our reported F1 scores are slightly lower than scores reported by previous research [2] using similar features and classifiers (F1 0.4 vs 0.5 for local and global features), however, that research was conducted using a short movie instead of MOOC lectures and free self-reporting instead of periodic self-reporting for obtaining mind-wandering reports.

With respect to the used classification methods, we find that the SMOTE method which is used to deal with unbalanced data can improve the performance of L2-Logistic Regression. However, it may have less improvement or negative impact on the performance of decision trees. This could be explained by that the decision tree is mainly based on the exact values of samples, while the SMOTE method generates "fake" samples for the minority class, thus diluting the result quality. The Gaussian Naive Bayes models outperforms the other methods on

Table 3: Statistics of detection results in leave-one-participant-out cross validation. $P_{highest}$ shows the detection results of the participant with highest f1-measure and P_{lowest} shows the detection results of the participant with lowest f1-measure.

Data	*Metrics*	*Max*	*Min*	*Mean*	*Std*	$P_{highest}$	P_{lowest}
Tobii	Precision	0.714	0	0.370	0.214	0.714	0
Data	Recall	1.000	0	0.571	0.319	0.833	0
	F1	0.769	0	0.381	0.203	0.769	0
WebGazer	Precision	0.667	0	0.341	0.198	0.600	0
Data	Recall	1.000	0	0.773	0.299	1.000	0
	F1	0.750	0	0.426	0.201	0.750	0

WebGazer data with each combination of feature sets including the SMOTE method.

The most interesting finding in this experiment is that we get higher *Recall* and *F1* scores based on features extracted by WebGazer. Based on our intuition, the detection results of Tobii data should be better than the detection results of WebGazer data, since the gaze data estimated by Tobii has a higher accuracy and a higher sampling frequency. A possible explanation is that the data aggregated from the low-frequency sampling and micro-saccade heuristic for the WebGazer data is on a higher abstraction level compared to Tobii data.

Based on above results in Table 2, in the following, we dig deeper into our experimental mind-wandering detection results. As mentioned in Section 2.2, we use leave-one-participant-out cross validation in our experiments. In order to know *whether or not detections can be made equally well for all participants*, we investigate the detection results on each participant separately. The methods with best results on Tobii data and WebGazer data in Table 2 are selected (again, full results are on [18]). We analyze the detection results on Tobii data with L2-Logistic Regression, Global features, and the SMOTE method. Similarly, the detection results on WebGazer data with Gaussian Naive Bayes, global and local features are also analyzed. The results are shown in Table 3.

Based on the results in Table 3, it is visible that the ranges of both precision and recall are about 0.7. The standard deviations of both precision and recall are about 0.2 to 0.3. Additionally, there is a clear difference between the participants with best and the worst detection results. Therefore, we conclude that *the detections cannot be made equally well for all participants in our experiments.* Further research is required to dig into the reasons behind this observation.

Based on the exploratory analysis in Section 3.1, we can find that mind-wandering is not evenly distributed throughout the runtime of a video. This leads to an interesting question: *"Can mind-wandering detections be made equally well across the entire length of the lecture videos?"*. To simplify the problem, we split

Table 4: Detection results on mind-wandering in different parts of the videos. *Part 1* means the first half part of the video, and *Part 2* means the second half part of the video.

		Solar Energy		*Nuclear Energy*	
Data	*Metrics*	*Part 1*	*Part 2*	*Part 1*	*Part 2*
Tobii	Precision	0.259	0.375	0.375	0.526
Data	Recall	0.538	0.500	0.600	0.588
	F1	0.350	0.429	0.462	0.556
WebGazer	Precision	0.458	0.276	0.455	0.438
Data	Recall	0.846	0.444	1.000	0.824
	F1	0.595	0.340	0.625	0.571

each video into two parts with same length. Then, for each part of the video, we use the data of the other part and the data of the other video to train the model and to detect the mind-wandering in the specific part of the video. The results are shown in Table 4.

Based on the results shown in Table 4, we can conclude that *the detection of mind-wandering cannot be made equally well across the entire length of the lecture videos in our experiments*. For Tobii data, we can find that the recall of the mind-wandering detection in different parts of the same video are similar, while the precisions of the second part of the video are both higher than those of the first past of the video. For WebGazer Data, the detection results on the second part of the video about solar energy are worse than the results on the first part, while the detection results of the video about nuclear energy show a similar trend. We assume that this is connected to our participants losing their focus towards the end of the study, however, a deeper future investigation is necessary to reliably explain the reasons for this behavior.

The last experiment in our work is based on the question *"whether or not a model trained on one video translates to good detections for another video?"*. In our previous experiments, the leave-one-participant-out cross validation shows that our method can detect a participant's mind-wandering based on the model trained by the gaze data and mind-wandering reports of other participants. In order to detect learners' mind-wandering in scalable fashion, we also need to know "to what extent we can detect learners' mind-wandering in one course based on the model trained in other courses". If we can obtain a good detection results for such scenarios, it means that there might be a general model which can be used in different lecture videos in large-scale studies (i.e., "train once, deploy everywhere"). In the following experiments, the detection results based on the training data and test data in the same course are used as the baseline. Then, we use data collected from one lecture video as training data and detect participants' mind-wandering in another lecture video. The experiment settings

Table 5: Detection results with model translation between two different MOOC videos. *From A* means the model trained by gaze data extracted from video A, and *To B* means in video B the pre-trained model is leveraged.

		From Solar		*From Nuclear*	
Data	*Metrics*	*To Solar*	*To Nuclear*	*To Nuclear*	*To Solar*
Tobii	Precision	0.306	0.429	0.318	0.314
Data	Recall	0.355	0.556	0.519	0.516
	F1	0.328	0.484	0.394	0.390
WebGazer	Precision	0.224	0.444	0.407	0.358
Data	Recall	0.484	0.889	0.815	0.613
	F1	0.306	0.593	0.543	0.452

for classifiers, feature sets and the SMOTE method on different kinds of data are same as in previous experiments. The detection results are shown in Table 5.

Based on the results in Table 5, the first finding is that the detection results with training data and test data from the same video have clear differences on different videos. It is visible that the method which use both training data and test data extracted from the video about nuclear energy outperforms the same method on the data extracted from the video about solar energy. If we compare the experiments with the same test data, we can find that the detection results based on the model trained on other videos are not lower than the detection results of the model trained on the same video. Therefore, we believe that *a model trained on one video can translates to good detections in other videos*, at least if the videos share similarities with respect to style and type as in our example.

4 Conclusions

In this paper, we present a novel technique for automatically detecting mind-wandering when watching MOOC lecture videos by exploiting gaze tracking data obtained using only simple and ubiquitously available webcams. Our approach relies on training a supervised machine learning model based on aggregated gaze fixation and gaze saccade features. The model is then able to detect mind-wandering for 30-60 seconds short video intervals. This complements previous research using specialized and expensive eye tracker hardware. In our experiments, we could show that our webcam-based approach is on par with using professional and expensive eye-trackers. This opens the way for large-scale experiments in real-world MOOC and online-learning settings, allowing for both investigating learners' mind-wandering behavior and the effectiveness of interventions based on mind-wandering detection in future research under realistic conditions.

We evaluated our approach with different feature subsets and classification algorithms, and could show F1 scores of 0.44 for eye-tracking-based gaze estimation and a comparable score of 0.48 for webcam-based estimation, both scores being significantly better than the baseline of using an frequency-adjusted guessing classifier (F1 score of 0.29). Also, our method is learner independent, i.e. we can train a model on one subset of of learners and use it to predict mind-wandering for a different set of learners, shown by our experiments with leave-one-out-cross-validation. Similarity, we could show that the model can be trained on one MOOC video, and can be used on a different MOOC video with only small losses in prediction performance.

Our work also shows some limitations as for example the small pool of participants all sharing similar educational backgrounds. Similarly, the number of evaluated MOOC videos is limited, and we chose on purpose MOOCs of a comparable (but very common) style. Thus, it is unclear how well our approach can be applied to completely different types of videos or user groups. Also, we relied on very simple (but established) features, and a performance increase is expected when more sophisticated features are introduced.

Another core contribution provided by our work is the openly published repository of data collected during our elaborately conducted controlled lab study, which can be obtained on our companion web-page [18]. In addition to including the mind-wandering reports of our experiment's participants, we also provide the full set of raw eye-tracking and gaze data obtained by the eye-tracker and webcam as well as the complete results of our data analysis. Access to that data allows for replicating our results as well as developing more sophisticated mind-wandering detection models based on more complex local or global feature sets, thus bolstering research activities in this relevant field by lowering the entry barrier and allowing for better comparability of different approaches.

References

1. R. Bixler and S. DMello. Toward fully automated person-independent detection of mind wandering. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 37–48. Springer, 2014.
2. R. Bixler and S. DMello. Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, 26(1):33–68, 2016.
3. D. M. Bunce, E. A. Flens, and K. Y. Neiles. How long can students pay attention in class? a study of student attention decline using clickers. *Journal of Chemical Education*, 87(12):1438–1443, 2010.
4. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
5. R. Engbert and R. Kliegl. Microsaccades uncover the orientation of covert attention. *Vision research*, 43(9):1035–1045, 2003.
6. M. A. Killingsworth and D. T. Gilbert. A wandering mind is an unhappy mind. *Science*, 330(6006):932–932, 2010.

7. K. K. Loh, B. Z. H. Tan, and S. W. H. Lim. Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies. *Computers in Human Behavior*, 63:943–947, 2016.
8. W. McKeachie and M. Svinicki. *McKeachie's teaching tips*. Cengage Learning, 2013.
9. R. McMillan, S. B. Kaufman, and J. L. Singer. Ode to positive constructive daydreaming. *Frontiers in psychology*, 4:626, 2013.
10. C. Mills, R. Bixler, X. Wang, and S. K. DMello. Automatic gaze-based detection of mind wandering during narrative film comprehension. June 2016.
11. A. Papoutsaki, N. Daskalova, P. Sangkloy, J. Huang, J. Laskey, and J. Hays. Webgazer: scalable webcam eye tracking using user interactions. IJCAI, 2016.
12. K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.
13. O. Rodriguez. The concept of openness behind c and x-moocs (massive open online courses). *Open Praxis*, 5(1):67–73, 2013.
14. D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.
15. J. Smallwood, D. J. Fishman, and J. W. Schooler. Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic bulletin & review*, 14(2):230–236, 2007.
16. T. Tan, H. Zou, C. Chen, and J. Luo. Mind wandering and the incubation effect in insight problem solving. *Creativity Research Journal*, 27(4):375–382, 2015.
17. K. Wilson and J. H. Korn. Attention during lectures: Beyond ten minutes. *Teaching of Psychology*, 34(2):85–89, 2007.
18. Y. Zhao, C. Lofi, and C. Hauff. Ectel 2017 companion webpage. `http://www.wis.ewi.tudelft.nl/ectel2017`, 2017.