RHAAPsody: RHEED Heuristic Adaptive Automation Platform Framework for Molecular Beam Epitaxy Synthesis

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Molecular beam epitaxy (MBE) is an atomically precise method for the synthesis of extremely thin films which may possess unique and desirable functionalities. The epitaxial growth process is typically monitored by reflection high energy electron diffraction (RHEED), presenting information on surface morphology, growth rate, and crystallinity. However, observing and interpreting RHEED patterns is both time intensive and complex. In this work, we are developing an artificial intelligence (AI)-driven pipeline to enable automatic monitoring of the deposition process via real-time RHEED image analysis (one image per second) for targeted materials. Our pipeline utilizes a pre-trained image model that encodes each RHEED pattern image into a feature vector. Changes in the RHEED pattern are detected via two analytics methods: a time series-based changepoint detection method that measures changes in pairwise cosine similarity between feature vectors, and a graph theoretic method that clusters feature vectors by cosine similarity. We implement the open source framework and detect physically meaningful changes in RHEED videos collected from the deposition of epitaxial thin films such as anatase TiO_2 on $SrTiO_3(001)$. We present the strengths and weaknesses of this approach and its potential use as the basis for on-the-fly feedback control of MBE deposition parameters.

1 Introduction

Epitaxial thin films are layers of crystalline materials grown on a single crystal substrate surface that play an important role not only in an industrial setting but also the development and study of materials with new and undiscovered properties. Synthesizing these films by molecular beam epitaxy (MBE) is a highly controllable, atomically precise method for producing novel, non-equilibrium, and metastable materials and composites with unique and targeted performance properties [3]. Traditionally, epitaxial growth is observed by reflection high energy electron diffraction (RHEED)[11], in which a high-energy electron beam is directed at the surface of the film at a shallow angle, producing a diffraction pattern on a phosphorescent screen or detector. The resulting diffraction pattern encodes information on surface morphology, growth rate, and growth dynamics through features like streaks and spots. During and after deposition, an expert qualitatively determines the features and quality of a film by observing and interpreting the RHEED patterns produced during growth. However, interpreting RHEED patterns is complex and subtle features or feature changes are missed even by observers with significant expertise. In many cases, once an undesired feature like surface roughness or secondary phase formation is unambiguously identified during the deposition, manual adjustments to deposition parameters are insufficient to reverse the outcome. This underscores the need for advanced RHEED pattern analysis to detect subtle changes before they become visible.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

Previously, offline-analysis has leveraged dimensionality reduction techniques to identify pattern changes [15, 17] however, these decomposition approaches heavily rely on consistent pixel position and in reality the geometries of a real system experience fluctuations. Existing techniques have also approached improvements in throughput analysis and instrument parameter space through Bayesian optimization [16, 7, 13]. Very simple closed loop control for RHEED analysis was demonstrated by Shen et al. [12] using a 3D ResNet to cycle temperature ramping either 'on' or 'off' specific to the formation of quantum dots. This approach utilized a training data set of 120 quantum dot deposition runs. Here, we present a rapid, material-agnostic approach for real-time RHEED pattern analysis that detects changes within a 1 second frame rate, ahead of visual identification. This method is demonstrated on pre-recorded data from a model epitaxial oxide thin film deposition and forms a basis for on-the-fly feedback control of MBE deposition parameters.

2 Methods

Our approach for real-time RHEED pattern analysis consists of three components: preprocessing, changepoint detection, and film surface descriptor identification. However, the descriptor identification, i.e. a UNet architecture for segmenting streaks and spots in RHEED images [9], is currently underdeveloped for the scope of this paper. In preprocessing, TIFF images are captured at 1 Hz, converted to 8-bit grayscale, cropped to 280x138 to exclude non-diffraction regions, resized, and standardized using wavelet denoising and histogram equalization. For changepoint detection, image features are extracted using a VGG16 model [10] pretrained on ImageNet [4], creating 512-dimensional feature vectors. Grayscale images are tiled into three color channels to match the model's input requirements. Images are compared using a cosine similarity kernel:

$$K(\mathbf{x}, \mathbf{y}) = \text{CosineSim}(\mathbf{x}, \mathbf{y}) = 1 - \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{|\mathbf{x}||\mathbf{y}|}. \tag{1}$$

The changepoint algorithm segments the time interval I into sub-intervals I_0 and I_1 where images within each sub-interval are similar, and between sub-intervals are dissimilar, identifying the changepoint time τ. To measure the dissimilarity within intervals, we employ a segmentation cost function commonly used by changepoint detection methods [5, 6, 14]:

$$\text{SegCost}(J) = |J| \left(1 - \frac{1}{|J|^2} \sum_{t,t' \in J} K(\mathbf{x_t}, \mathbf{x_{t'}}) \right), \tag{2}$$

We detect changepoints in the RHEED video by optimizing the choice of τ using the segmentation cost of the full time interval compared against the segmentation cost of the sub-intervals separated by τ,

$$\tau = \arg\max_{\tau' \in I} \left\{ \frac{\text{SegCost}(I) - \text{SegCost}(I_0) - \text{SegCost}(I_1)}{|I|} \right\}. \tag{3}$$

If the maximum value in equation 3 exceeds the threshold ($h = 0.05$), τ is declared a changepoint. A schematic of the changepoint detection process is shown in Appendix A.1. To support online changepoint detection and reduce computational cost, we limit the width of I from the most recent frame to the closest frame of either: the most recent changepoint, the beginning of the video, or a sliding window width of 300 seconds.

The similarity matrix is also used for graph-based clustering, forming graph $\mathbf{G_t}$ with nodes representing images and edge weights corresponding to cosine similarity. Singular value decomposition is performed on $\mathbf{G_t}$ and nodes are clustered and visualized by the graph theoretic singular values [1]. Stability of clusters is measured by the ratio of the first two singular values [2]. If the stability ratio produces an inflection point when measured against time, then a changepoint is recorded.

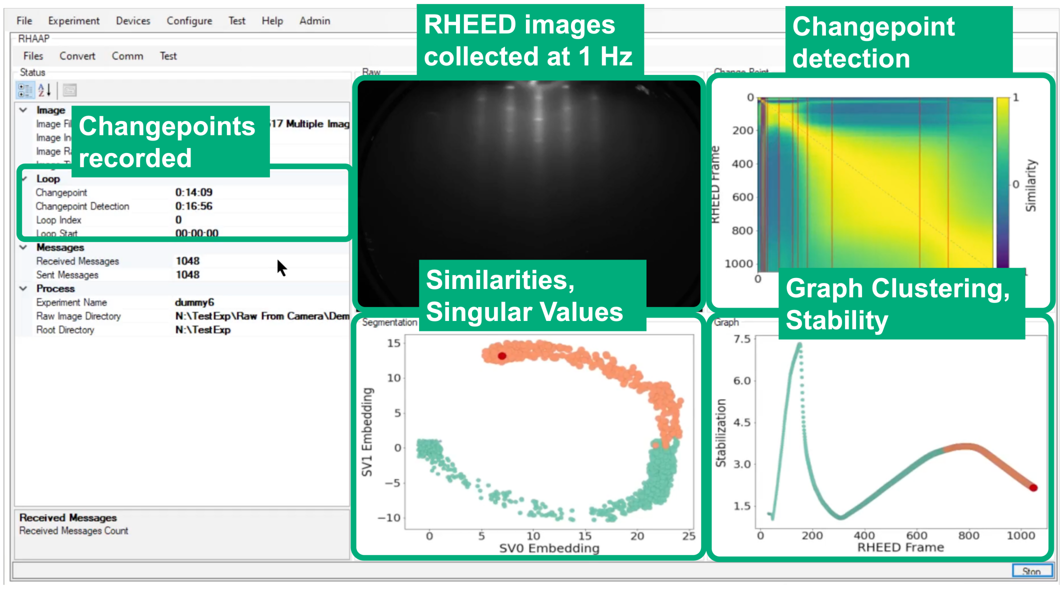
3 Implementation and Availability

3.1 Experimental Data

We prototyped the implementation of this framework on an anatase TiO2 film growth on a SrTiO3(001) substrate, chosen as a model thin film system because the stoichiometry is simple, which reduces the number of variables, but diverse outcomes can be expected under different deposition conditions: smooth vs. rough anatase, epitaxial anatase vs. nucleation of epitaxial rutile, polycrystalline rutile, amorphous TiO2. We focus on the transition from an epitaxial film with a smooth surface to an epitaxial film with a rough, islanded surface. This transition is typically observed in RHEED diffraction patterns as a change from diffraction streaks to diffraction spots, with the spots lying along the streak positions. The code for changepoint detection, graph analytics, and instrument communications are available at `https://github.com/anonymized_for_review`.[1]

3.2 On-the-fly Communication Framework

RHEED images are collected in raw format from the instrument at 1 Hz to a shared folder on the CPU controller computer. The images are then converted to TIFF format with a header and distributed to a lambda GPU computer (see Appendix A.3) through a ZeroMQ publish-subscribe model [8] to be analyzed in real-time.[2] Each image feeds directly into the pre-trained deep convolutional neural network (VGG16) to be converted into 512-dimensional feature vector. The feature vector is compared against previous images with the statistical analysis of changepoints. Additionally, the image feature vectors expand a sliding window and the network graph analysis determines change in clustering of new image(s) as described in Section 2. Analytics from the changepoint and graph clustering methods are returned through the subscriber of the ZeroMQ framework and rendered in a display interface shown in Figure 1.



Changepoint 0:14:09
Changepoint Detection 0:16:56
Loop Index 0
Loop Start 00:00:00
Received Messages 1048
Sent Messages 1048
Experiment Name dummy6
Raw Image Directory N:\TestExp\Raw From Camera\Dem
Root Directory N:\TestExp

Figure 1: Visualization of RHAAPsody interface highlighting changepoint detection and graph clustering analytics displayed through the ZeroMQ framework. This preliminary framework demonstrates the message exchange process and lays the groundwork for automated instrument control once changepoints are correlated with directional changes, i.e. increase or decrease, in instrument parameters.

Currently, this interface is used to communicate experimental metadata, display data collection and analytics, and indicate when instrument control could intervene. The next phases of our research will incorporate directional changes in instrument parameters, i.e. temperature, in response to detected changepoints.

---

[1] This link is under disclosure review and will be available on or after September 12, 2024.

[2] The prototyping experiment was prerecorded and simulated as real-time in order to leverage existing growth data.

3.3 Results

The prototype TiO2 film growth was analyzed with changepoint detection and graph theoretic “stabilization” in the ZeroMQ framework as well as an after-the-fact video analysis from an MBE subject matter expert. The corresponding timeline for identified changes is shown in Figure 2. The change detected, barely visible in the 720s RHEED image, is the appearance of small dots along the 3 center-most streaks.

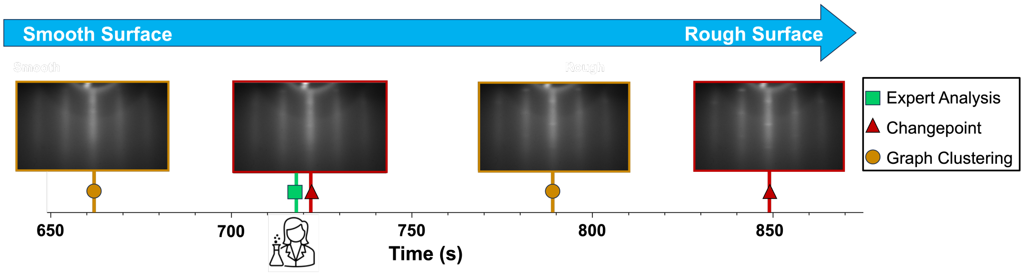


Figure 2: Comparison of expert analysis, changepoint detection, and graph theoretic identified ‘changes’ during the anatase TiO2 film growth on a SrTiO3(001) substrate. The identified change is the faint appearance of spots occurring along the three center-most streaks, illustrating the transition between smooth and rough surfaces over time.

4 Discussion

4.1 Conclusions

All the ML methods implemented in RHAAPsody—VGG-16, changepoint detection, and graph clustering— are executed within the targeted 1 Hz acquisition rate, making our tool suitable for real-time analysis during thin film deposition. Both the changepoint detection and the graph based clustering methods indicate accurate changing growth characteristics, such as the pattern transitioning from streaky to spotty, with the graph clustering method identifying the change sooner. The graph clustering correctly separates streaky images from spotty images, separating images which indicate layer-by-layer growth from amorphous growth patterns. Additionally, both the changepoint detection and graph clustering methods are material-agnostic and operate without supervision. Finally, our methodology is less sensitive to pixel scale and translation compared to matrix factorization techniques like Principal Component Analysis.

4.2 Limitations

The results in Section 3 describe a single dataset. We continue to collect data to both develop our communication framework as well as our methods, especially as we begin to incorporate instrument parameters into our analysis (e.g. temperature and partial pressure of Oxygen). Preliminary results on new experiments show promise as seen in Appendix A.2. While our methods function in real-time, because τ is optimized over the entire time interval a changepoint is identified only after it has already occurred, making the process retrospective. Both changepoint detection and graph analytics may be affected by noise in the initial seconds of an experiment. Lastly, neither approach is predictive; therefore, reversing an undesired outcome, such as a trend towards spotty or rutile film, is not feasible. Our current and future work aims to integrate changepoint detection and graph analytics with predictive models like convolutional LSTMs to identify changes early and enable real-time feedback control. We are continuously collecting data to better understand the parameter space around growth and instrument control. The ZeroMQ framework is designed to facilitate both the relay and control of instrument parameter settings.

References

[1] H. Abdi. Singular value decomposition (svd) and generalized singular value decomposition. *Encyclopedia of measurement and statistics*, 907(912):44, 2007.

[2] E. Andreotti, D. Edelmann, N. Guglielmi, and C. Lubich. Measuring the stability of spectral clustering. *Linear Algebra and its Applications*, 610:673–697, 2021.

[3] S. A. Chambers. Epitaxial growth and properties of doped transition metal and complex oxide films. *Advanced Materials*, 22(2):219–248, 2010.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

[5] D. Garreau and S. Arlot. Consistent change-point detection with kernels, 2017. URL `https://arxiv.org/abs/1612.04740`.

[6] Z. Harchaoui and O. Cappe. Retrospective mutiple change-point estimation with kernels. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 768–772, 2007. doi: 10.1109/SSP.2007.4301363.

[7] S. B. Harris, A. Biswas, S. J. Yun, K. M. Roccapriore, C. M. Rouleau, A. A. Puretzky, R. K. Vasudevan, D. B. Geohegan, and K. Xiao. Autonomous synthesis of thin film materials with pulsed laser deposition enabled by in situ spectroscopy and automation. *Small Methods*, page 2301763, 2024.

[8] P. Hintjens. *ZeroMQ: Messaging for Many Applications*. O'Reilly Media, 2013.

[9] H. Liang, V. Stanev, A. G. Kusne, Y. Tsukahara, K. Ito, R. Takahashi, M. Lippmaa, and I. Takeuchi. Application of machine learning to reflection high-energy electron diffraction images for automated structural phase mapping. *Phys. Rev. Mater.*, 6:063805, Jun 2022. doi: 10.1103/PhysRevMaterials.6.063805. URL `https://link.aps.org/doi/10.1103/PhysRevMaterials.6.063805`.

[10] S. Liu and W. Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian conference on pattern recognition (ACPR)*, pages 730–734. IEEE, 2015.

[11] J. E. Mahan, K. M. Geib, G. Robinson, and R. G. Long. A review of the geometrical fundamentals of reflection high-energy electron diffraction with application to silicon surfaces. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 8(5):3692–3700, 1990.

[12] C. Shen, W. Zhan, K. Xin, M. Li, Z. Sun, H. Cong, C. Xu, J. Tang, Z. Wu, B. Xu, et al. Machine-learning-assisted and real-time-feedback-controlled growth of inas/gaas quantum dots. *Nature Communications*, 15 (1):2724, 2024.

[13] N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature*, 624(7990):86–91, 2023.

[14] C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2019.107299. URL `https://www.sciencedirect.com/science/article/pii/S0165168419303494`.

[15] R. K. Vasudevan, A. Tselev, A. P. Baddorf, and S. V. Kalinin. Big-data reflection high energy electron diffraction analysis for understanding epitaxial film growth processes. *ACS nano*, 8(10):10899–10908, 2014.

[16] Y. K. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, and H. Yamamoto. Machine-learning-assisted thin-film growth: Bayesian optimization in molecular beam epitaxy of srruo3 thin films. *APL Materials*, 7(10), 2019.

[17] A. Yoshinari, Y. Iwasaki, M. Kotsugi, S. Sato, and N. Nagamura. Skill-agnostic analysis of reflection high-energy electron diffraction patterns for si (111) surface superstructures using machine learning. *Science and Technology of Advanced Materials: Methods*, 2(1):162–174, 2022.

A Supplemental material

A.1 Changepoint Schematic

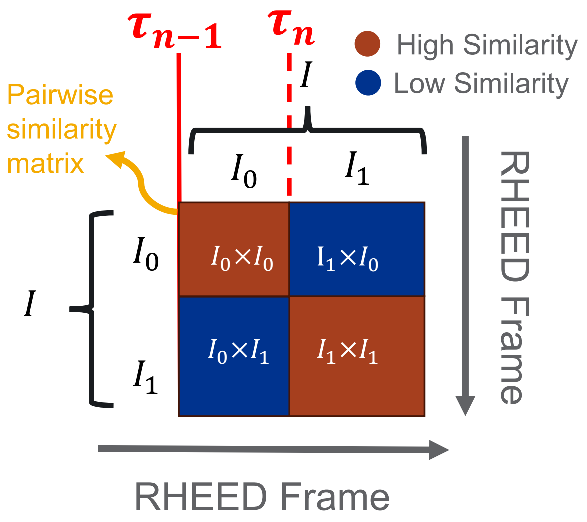


Figure 3: A schematic of the changepoint detection process. The time interval between the last changepoint and the current time is segmented such that the diagonal blocks of the similarity matrix (internal similarity scores) have high overall values, while the off diagonal blocks (cross similarity scores) have low values.

A.2 Additional Experimental Results

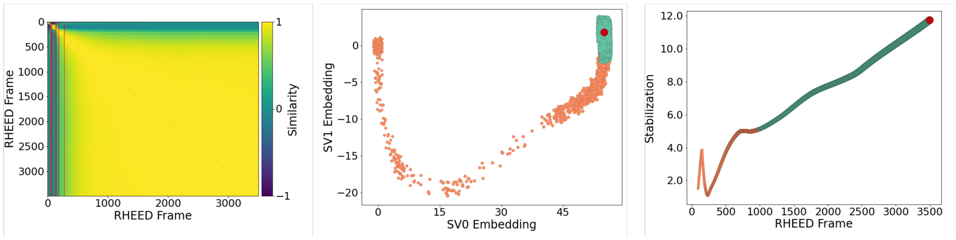


Figure 4: Changepoint and graph analytics on a secondary experiment where no changes in film growth occur after the initial growth phases. Initial growth is indicated by multiple detected changepoints early on.

A.3 GPU Specifications

The Lambda Vector GPU Workstation used in this experiment was purchased directly from Lambda in July 2023. The workstation has the following specifications:

- **Operating system**: Ubuntu 22.04, includes Lambda Stack for managing TensorFlow, PyTorch, CUDA, cuDNN, etc.
- **Processor**: AMD Ryzen Threadripper PRO 5955WX: 16 cores, 4.0 4.5GHz, 64 MB cache, PCIe 4.0
- **GPU**: 2x NVIDIA RTX A6000: 48GB memory, 10752 CUDA cores, 336 Tensor cores, NVLink
- **System memory**: 256 GB: DDR4-3200 UDIMM

- **OS drive**: 1x 1.92 TB M.2 NVMe
- **Data drive**: 2x 15.36 TB U.2 NVMe: Data center SSD, 1 DWPD, PCIe 4.0
- **Onboard networking**: 2x 10 Gbps RJ45 Ethernet ports, 1x dedicated IPMI port

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We in fact describe the implementation of an open source framework to detect physically meaningful changes in RHEED videos collected from the deposition of anatase TiO2 on SrTiO3(001). We present the strengths and weaknesses of our methods in Section 4 as well as potential use as the basis for on-the-fly feedback control of MBE deposition parameters.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are described in Section 4.2, centering around data limitations, factors that may affect performance (experiments are cost and labor intensive, we continue to grow in this area) and timeliness of our algorithms. Computational efficiency is described as a non-issue given the low latency of the RHEED collection framework.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: The paper is an applied technique and does not include any theoretical proofs.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The code, data, and implementation described are available on `https://github.com/anonymized_for_review`.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code, data, and implementation described are available on `https://github.com/anonymized_for_review`. The streaming framework is available in code and visually represented in Appendix A.1.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The standing implementation of our approach does not include training/test experiments. Future work will leverage collected data in an experimental setting, but is too early in concept and data volumes to describe.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The limitations described in Section 4.2 reference the demonstration of a single film growth, which is not enough to qualify the statistical significance of our approach. As more data is collected and annotated, we will understand statistical significance in context of subject matter expertise for event detection. To our knowledge, no annotated and available RHEED pattern dataset currently exists.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: These methods are not computationally intensive, however compute is loosely described (CPU and GPU) in the setting of on-the-fly analysis.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research conforms with and appreciates the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

Justification: The work performed is foundational research and not directly tied to societal impacts. We do describe the importance of an analysis of this time to materials settings alone.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This work does not pose a high risk for misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Existing models and data are properly cited at a manuscript level.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The open sourced code provided for this work includes licensing and use descriptors. The acknowledgement statement includes asset source descriptors with use outlined in the open source code made available.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: While a subject matter expert annotated 'ground truth' in our experiment, the paper does not include crowdsourcing nor human subjects research.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

Justification: While a subject matter expert annotated ‘ground truth’ in our experiment, the paper does not include crowdsourcing nor human subjects research.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.