Data Engineering and Applications

Research Paper

Edge-Ready Intelligent Monitoring of Batch Crystallization Processes Using Hierarchical Image Segmentation and Data-Driven Endpoint Analytics

Hrvoje Petković[1, *], Ante Španjol[1] and Domagoj Mihalić[2]

[1] Faculty of Information Technology, Virovitica University of Applied Sciences, Virovitica, 33000, Croatia
[2] Faculty of Information Technology, University of Split, Split, 21000, Croatia
*Corresponding author: hrvoje.pet@virovitica.hr

Abstract. In this study, a Hiera-UNet model based on online particle pictures is proposed for endpoint recognition in batch crystallisation. Because of the constant changes in crystal size, boundary clarity, aggregation, blur, lighting, bubbles, and mother-liquid backdrop during the batch, endpoint detection is challenging. The suggested system reconstructs particle masks, generates overlapping crystal-region patches, extracts hierarchical morphological characteristics, normalises online particle frames, and transforms segmentation evidence into temporally smoothed endpoint probability. Particle masks, morphological labels, illumination disturbance markers, and endpoint windows are supplied, and a simulated picture stream including 12,600 frames of nucleation, growth, aggregation, near-endpoint, and post-endpoint phases is utilised for assessment. Hiera-UNet enhanced endpoint accuracy from 87.2% to 94.6%, decreased average endpoint-delay latency from 18.4 s to 7.9 s, and raised the mask Dice score from 0.841 to 0.906 when compared to U-Net, DeepLabv3+, SegFormer, and CNN-LSTM baselines. The live monitoring of crystallisation end-points can be supported by hierarchical segmentation, boundary-aware morphology extraction, and decision smoothing based on the results; however, recipe-specific calibration and validation using actual probe-image streams are still required.

Keywords: *Online Particle Image Recognition, Hiera-UNet, Batch Crystallization Endpoint, Crystal Morphology Segmentation, Boundary-aware Decoder, Process Vision Monitoring*

Received on 04 April 2023, Accepted on 09 August 2023, Published on 17 August 2023

Introduction

During batch crystallisation, nucleation, growth, aggregation, dissolution, and shape changes may be seen by online particle and picture monitoring. Deep learning can learn visual patterns that are challenging to express with fixed monitoring rules, as demonstrated by LeCun et al. [1], and general computer-vision research [2] suggests that representation learning lessens reliance on manually created descriptors when object size and texture change concurrently. Because the goal state is not a single object category, endpoint identification is more challenging than generic picture categorisation. Aggregation, background clarity, average size, shape stability, and particle boundary density all fluctuate simultaneously. Additionally, local visual evidence is not accessible when targets are close together, as in slurry photos from online probes, according to research on object detection [3]. As a result, the end point model ought to link process-level stage interpretations with pixel-level crystal data.

The end of crystallisation is often determined by offline sampling, turbidity curves, focused-beam reflectance trends, and operator expertise. Although the techniques are workable, they could react slowly to slight variations in the live particle-image stream. Robust visual models should be able to manage illumination, blur, and form fluctuations, which are typical in online crystallisation pictures with bubbles, partial occlusion, probe-window reflection, and uneven lighting, according to image augmentation research [4]. Scale variation should be addressed early on, according to reviews of deep visual models for applied imaging [5]. A little shift in visuals might be the result of typical variations since process pictures and industrial anomaly detection share certain commonalities [6]. Reporting endpoint probability and morphology-based evidence is also supported by Explainable AI research [7].

Another issue is data management. Online particle imaging is compatible with machine-learning data cleaning studies and can demonstrate how stream preprocessing impacts downstream dependability [8]. Particle size will be artificially reduced in a low-exposure picture, and the crystal edges needed for endpoint detection may be hidden in a blurred frame. Vision Transformer studies [10] have demonstrated that hierarchical context may link local and global visual information, and computer-vision attention research [9] implies that relevant local regions should be given greater weight than uniform backdrop areas. These concepts underpin a Hierarchical UNet architecture that converts online particle pictures into temporally stable endpoint-confidence curves by cleaning, normalising, segmenting, and transforming them.

This is how the remainder of the paper is structured. In Section 2, hierarchical vision techniques, crystal segmentation models, and online particle image monitoring are demonstrated. Image-stream creation, Hiera-UNet multi-scale feature fusion, and endpoint probability calibration are demonstrated in Section 3. The simulated data, segmentation outcomes, endpoint-recognition performance, ablation analysis, robustness test, and deployment interpretation are all displayed in Section 4. The method's primary contents and applications are summarised in Section 5.

Related Work

Online Particle Image Monitoring in Batch Crystallization

Particle population changes, crystal edge development, agglomeration propensity, and slurry-background fluctuations may all be immediately seen by online particle imaging. Encoder-decoder segmentation [11] demonstrates that low-level border recovery and high-level semantic context should be coupled, and self-configuring segmentation work [12] offers a useful reference for adjusting the model structure to image attributes. This connection is necessary because a little bright spot in the crystallisation picture might represent a reflection from the probing window, a noise point, or a newly formed crystal. As a result, the monitoring job should simultaneously have frame-level consistency and local contour sensitivity.

While dense feature reuse [13] shows how intermediate visual patterns may be recycled across network depth, nested skip-connection models [14] have enhanced fine-structure preservation in segmentation. Since endpoint evidence typically emerges gradually, these concepts are appropriate for crystal picture streams. It is possible for the same process to produce huge agglomerates in one location and small particles in another. The model will be unable to identify the boundary instability end point if it just takes into account the mean particle size. As a result, a process-image model should concurrently display population density, edge sharpness, shape distribution, and temporal stability.

Crystal Morphology Segmentation and Endpoint Recognition

The figure will display the degree of crystallisation, allowing for endpoint identification and morphological segmentation. Schlemper et al. [15] have demonstrated that Selective Focus can improve boundary-sensitive segmentation performance, and MultiResUNet research [16] has also found that local feature reuse helps irregular target objects. Agglomerated crystals frequently overlap and can be as big as individual massive particles. As a result, the segmentation masks should be viewed as proof of an intermediate step rather than the final findings.

Two approaches to reliable visual representation are residual feature learning [17] and effective backbone search [18]. Because the late-stage changes are tiny and may only affect a portion of the image, crystallisation endpoint detection is different from general image categorisation. Multi-scale context can aid in integrating local boundary information with global image-stage information, according to token-based visual representations [19] and hierarchical transformer-based visual models [20]. In order to integrate segmentation, morphology extraction, and endpoint smoothing into a single online recognition framework, the Hiera-UNet architecture was selected in this study.

Methodology

Online Particle Image Stream Construction

First, the online particle-image stream is used to create normalised time-window samples. The acquisition time, weak endpoint label, estimated slurry-density state, probe-illumination status, and morphology-stage annotation are all associated with each frame. After the input image is transformed into a normalized-intensity frame, crystal-edge evidence may be compared under various lighting scenarios. Figure 1 depicts the whole online endpoint recognition procedure for particle image capture and decision output. The process from probe-image input to normalisation, patch construction, Hiera-UNet segmentation, morphological feature extraction, endpoint probability, temporal smoothing, and operator-facing output is shown in Figure 1.

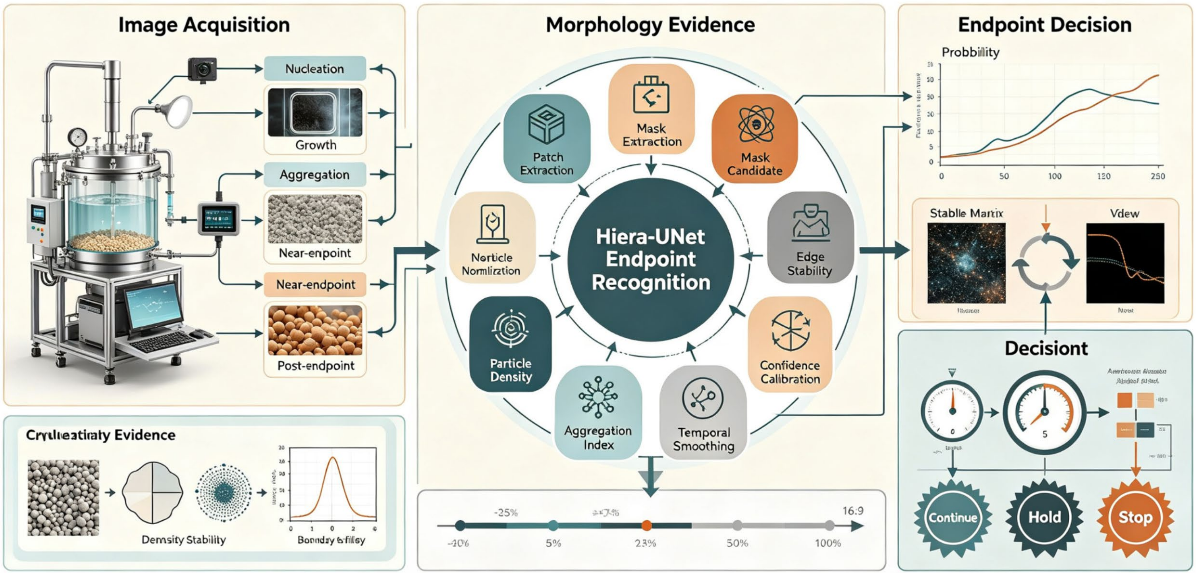

Figure 1. Online particle image endpoint-recognition workflow.

$$I'_t = \frac{I_t - \mu_t}{\sigma_t + \epsilon} \qquad \text{Eq.(1)}$$

Here, I_t denotes the original online particle frame, while the normalized frame I'_t reduces the influence of local exposure variation. The small constant prevents numerical instability when the frame is nearly uniform. The normalized image is then divided into overlapping patches so that small nuclei and large agglomerates can be observed within the same time window.

$$X_t = \{P_{t,1}, P_{t,2}, \dots, P_{t,n}\} \qquad \text{Eq.(2)}$$

The patch set X_t keeps local particle structures while preserving the temporal identity of the frame. This representation is useful because endpoint evidence is not uniformly distributed. Some regions may already show stable crystals, while other regions still contain evolving particles or blurred edges. The construction therefore gives the later model access to both local details and stream-level continuity.

Hiera-UNet Multi-scale Feature Fusion

Crystal morphology is extracted using a number of encoders and decoders at various scales. A fine nucleus and edge detail can be seen at shallow magnification; local clustering and some contour blurring can be seen at medium magnification; crystal-population order and the background mother-liquid may be seen at high magnification. Figure 2 displays the endpoint feature-fusion structure and Hiera-UNet segmentation. Skip fusion combines fine and coarse details, as seen in Figure 2, and the boundary-aware reaction indicates if the crystal shape is evolving. Particle masks are not the ultimate product; rather, the segmentation quality is linked to the process-stage recognition.

https://doi.org/10.64972/dea.2023.v2i3.3715d:52-64

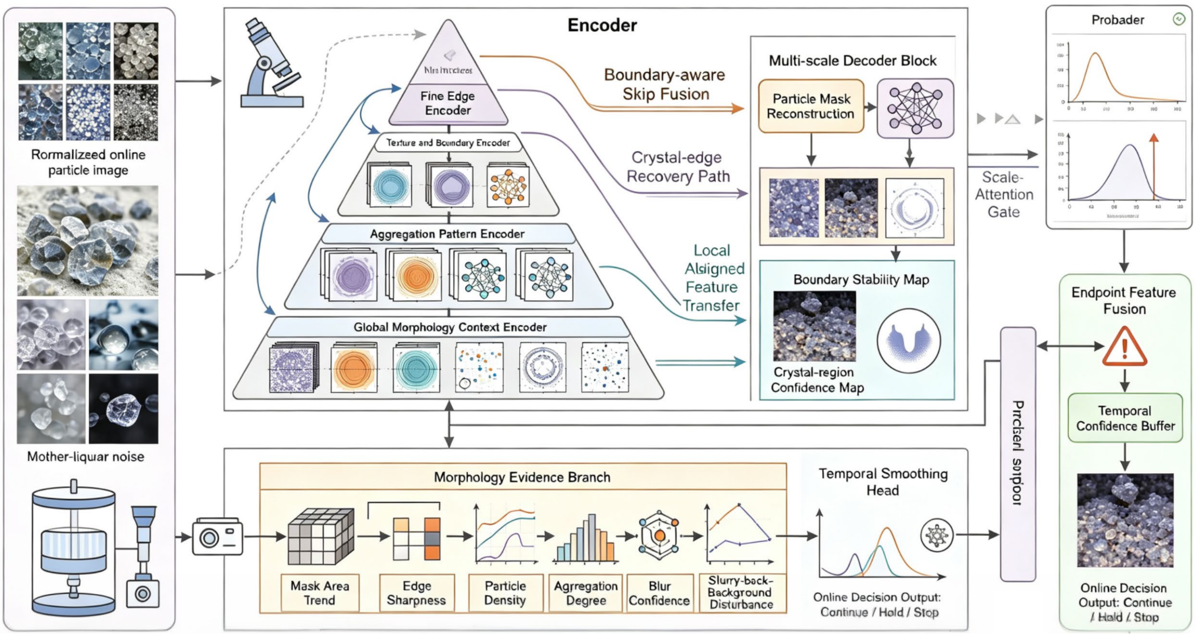

Figure 2. Hiera-UNet multi-scale segmentation and feature-fusion structure.

$$F_l = \phi_l(F_{l-1}, I_t') \quad \text{Eq.(3)}$$

The feature map F_l represents the response of the lth hierarchical stage. The mapping function includes convolutional extraction, normalization, and local attention. Low-level maps are sensitive to crystal edges, while high-level maps describe morphology layout and crystal population density.

$$\hat{M}_t = \psi(F_1, F_2, \dots, F_L) \quad \text{Eq.(4)}$$

The decoder function fuses multi-scale features into the particle mask. Instead of depending only on the deepest feature map, the model uses skip fusion to recover blurred boundaries. This is important because the boundary between crystal and mother liquor is often weak near the endpoint.

$$z_t = GAP([F_L, \hat{M}_t, B_t]) \quad \text{Eq.(5)}$$

The vector z_t aggregates deep morphology features, predicted particle regions, and boundary cues. This vector is later used for endpoint recognition. It is deliberately built from both segmentation and image-level features, because segmentation quality alone cannot fully describe process completion.

Endpoint Probability Calibration and Decision Rule

An instantaneous endpoint probability of the morphological feature vector is produced by the endpoint recogniser. Because of light flicker, probing disturbance, bubble motion, or brief mixing that may be misinterpreted for a steady picture, a single frame is not used as the final result. As a result, the probability is smoothed over a brief period of time, and the endpoint label is only given when the confidence for successive frames stays over a level chosen for validation. This rule preserves the hierarchical visual evidence produced by Hiera-UNet, retains a reasonably basic model appropriate for online deployment, and strikes a balance between early endpoint detection and delayed detection.

$$p_t = \sigma(W z_t + b) \quad \text{Eq.(6)}$$

The value p_t is the instantaneous endpoint confidence. It increases when the crystal mask becomes stable, the boundary signal becomes less variable, and the morphology feature vector moves toward the learned near-endpoint state.

$$\bar{p}_t = \alpha p_t + (1 - \alpha)\bar{p}_{t-1} \quad \text{Eq.(7)}$$

The smoothed confidence reduces false alarms in a noisy image stream. A high value must persist for several consecutive frames before the batch is classified as reaching the endpoint.

https://doi.org/10.64972/dea.2023.v2i3.3715d:52-64

$$y_t = 1(\bar{p}_t > \tau) \quad \text{Eq.(8)}$$

The decision rule produces the final endpoint label. The threshold is selected on the validation set to balance early warning and delayed recognition. This rule keeps the model simple enough for online deployment while retaining the hierarchical visual evidence produced by Hiera-UNet.

Experiments and Discussion

Dataset Setting and Crystal Morphology Distribution

The simulated online crystallization image dataset contains 12,600 frames from nucleation, early growth, steady growth, near-endpoint, and post-endpoint stages. Real-world industrial anomaly data [21] show the need to address difficult visual cases because normal variations can resemble process faults, and transformer detection research [22] shows that dense image reasoning benefits from global context when fine nuclei, developing crystals, and agglomerates appear in one frame. The dataset includes particle masks, endpoint labels, illumination-disturbance markers, blur categories, and morphology-stage annotations. The dataset configuration and morphological distribution are shown in Figure 3. Figure 3(a) shows that endpoint windows are rarer than typical growth-stage frames. Figure 3(b) shows particle-size intervals, Figure 3(c) shows blur-level categories, and Figure 3(d) shows endpoint-window density across the batch.

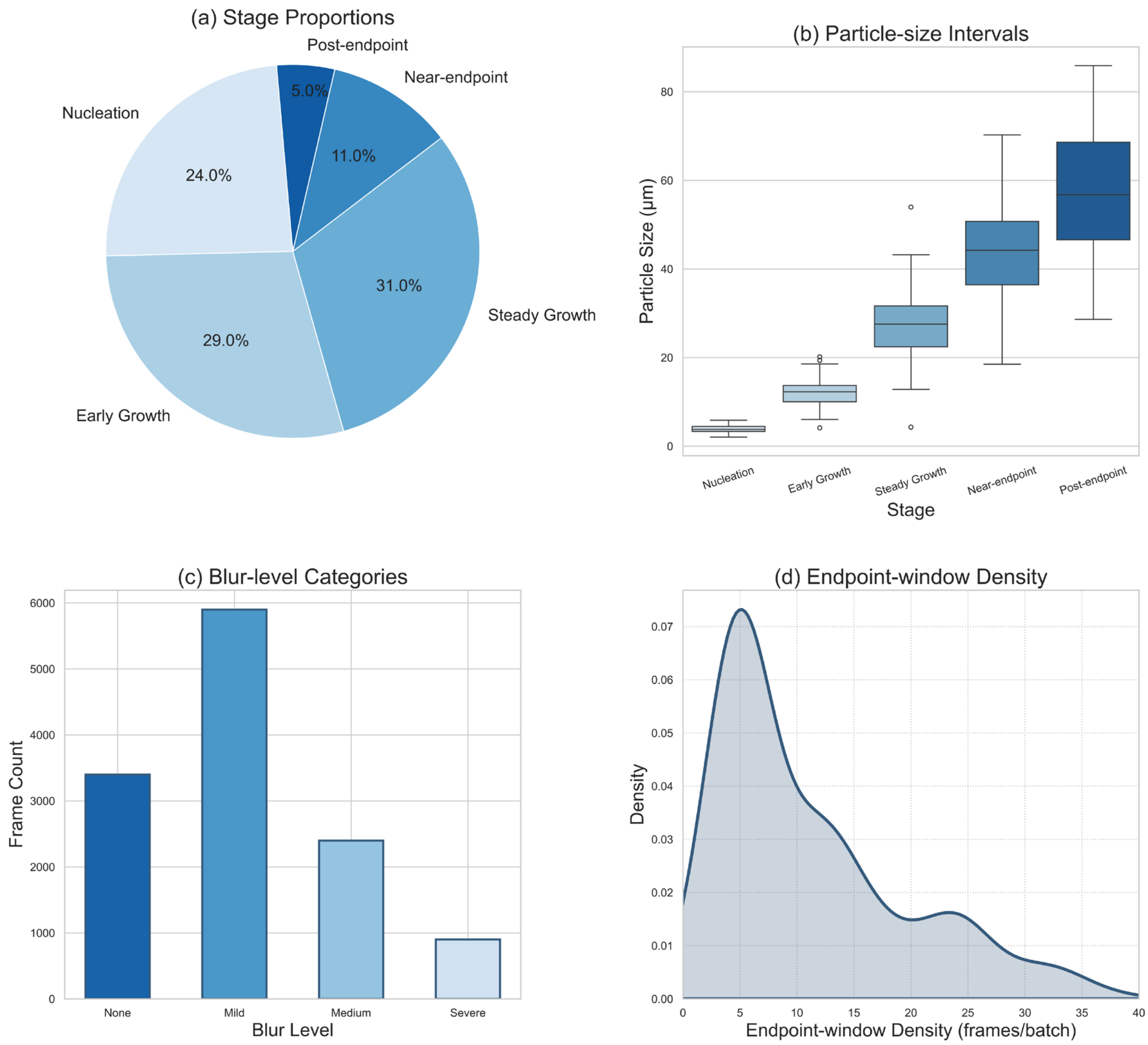

Figure 3. Dataset setting and crystal morphology distribution. (a) Stage proportions. (b) Particle-size intervals. (c) Blur-level categories. (d) Endpoint-window density.

https://doi.org/10.64972/dea.2023.v2i3.3715d:52-64

As shown in the above pattern, endpoint recognition is not the same as large-crystal detection. Dense loss design [23] proposes that difficult foreground samples should be given more weight, which is suitable for rare endpoint windows and unclear particle boundaries. Early growth frames in the simulated data have many fine particles but weak endpoint evidence, and near-endpoint frames may still contain local mixing disturbances. Image restoration research [24] aims to retain blur-level information so as not to affect subsequent predictions, and dehazing studies [25] have shown that background degradation reduces visual reliability. Therefore, the dataset is designed to determine whether a stable particle mask, boundary consistency and morphology confidence occur simultaneously near the endpoint.

Sequence-aware division is used in the dataset split. Although neighbouring time frames from the same simulated batch are not randomly allocated between subsets, 70% of the batches are used for training, 10% for validation, and 20% for testing. This is not temporal leaking; rather, it happens when a model does well just because it was trained on almost identical frames. Only the test set is utilised to present the final segmentation and endpoint-recognition results; the validation set is used to choose the endpoint threshold and smoothing coefficient. Because a deployed system must handle a fresh batch rather than duplicate an earlier one, this mode is more similar to practical batch operation.

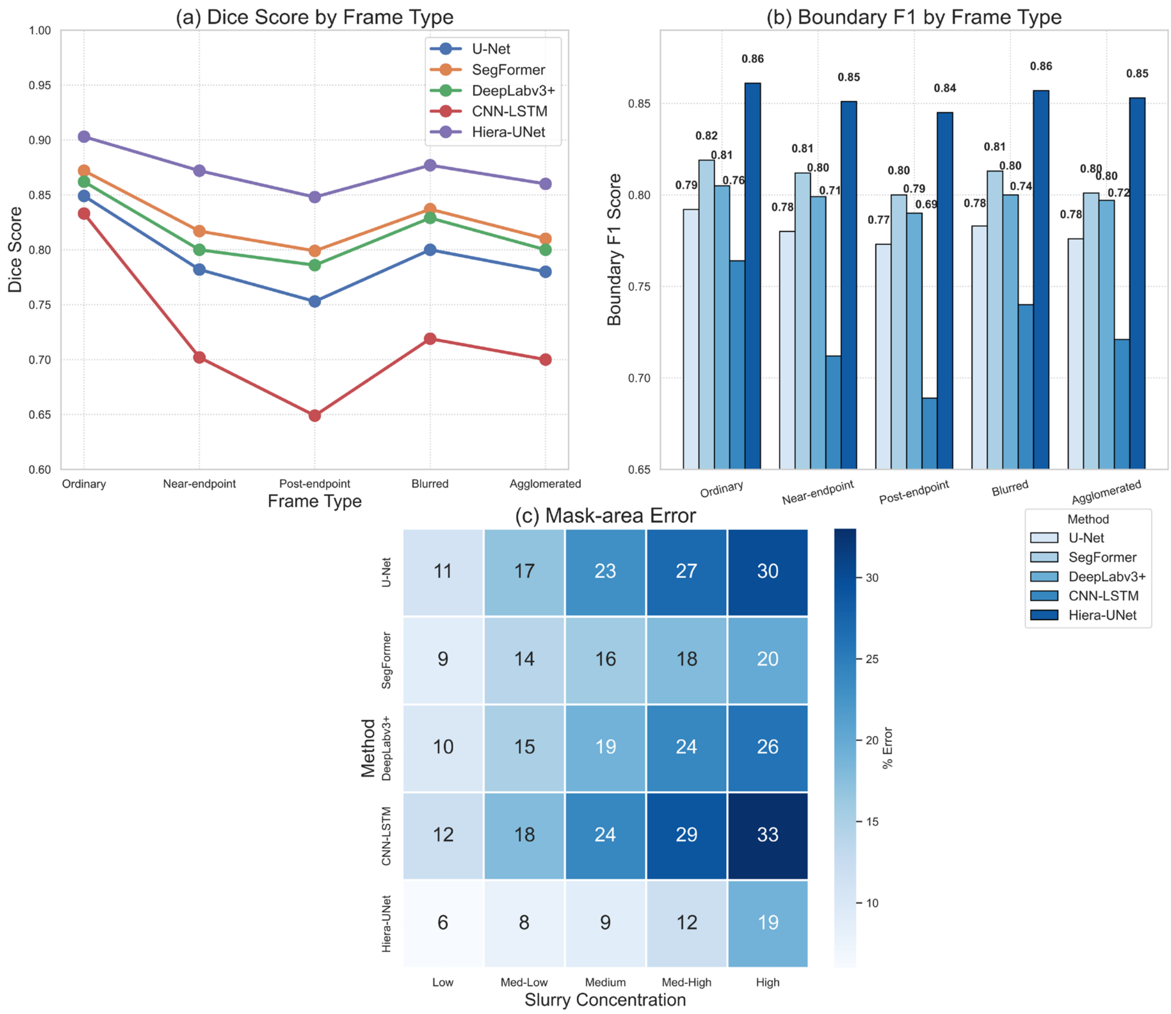

Figure 4. Segmentation accuracy and boundary recovery. (a) Dice score. (b) Boundary F1. (c) Mask-area error.

How the results are interpreted is also influenced by the distribution of shapes. When a near-endpoint picture has tiny bubbles, slurry streaks, or overlapping crystals, a model that works well on clear mid-growth images could still fail. As a result, the dataset now includes both visually deteriorated and clean frames. The endpoint label is not based on a particular mask region, but rather on the stability of morphological evidence over a brief period of time. Although the design will be more useful for online control, it is somewhat challenging to replicate.

It ascertains if the model has acknowledged that the crystal population is sufficiently stable to justify an end-point choice. Even if the visible particle population's appearance varies within the same batch, a decent endpoint-recognition model should still function effectively on the dominating morphological class. Plate-like crystals are wide reflectors, needle-like crystals have long, thin borders, and clustered crystals are challenging to separate separately. Agglomerated regions are especially troublesome because, although having an unequal internal size distribution, they may appear to have completed endpoints. As a result, the figure raises the bar for model performance: great precision should only be attained when applied to a wide range of forms, densities, light conditions, and manufacturing phases. As a result, the simulated labels are arranged based on endpoint states as well as picture classifications. The study also clarifies why the border F1, mask-area error, endpoint delay, and online warning curves are reported together in the subsequent experiments. Each index prevents the same accuracy figure from being repeated and represents a specific shortcoming in the dataset.

Endpoint-recognition Performance and Segmentation Comparison

The CNN-LSTM endpoint classifier, U-Net, U-Net++, DeepLabv3+, and SegFormer are compared with the suggested model. Figure 4 displays boundary recovery and segmentation accuracy. The prediction's Dice coefficient for general and near-endpoint frames in relation to the reference mask is displayed in Figure 4(a). The border F1, seen in Figure 4(b), is more susceptible to fuzzy crystal edges and overlapping particles. The mask-area inaccuracy at various slurry concentrations is shown in Figure 4(c), and incorrect region expansion often happens. Multi-level visual evidence is advantageous for dense prediction [26], attention blocks can concentrate on informative crystal areas [27], and research on panoptic segmentation [28] demonstrates that pixel-level categorisation is not adequate when both object borders and region identities are needed. The Dice coefficients for Hiera-UNet, U-Net, and SegFormer are 0.906, 0.841, and 0.873, respectively; the border F1 score rose from 0.792 to 0.861.

Endpoint recognition performance is shown in Figure 5. Figure 5(a) compares endpoint accuracy and clarifies whether the improvement reflects a better process decision rather than only a cleaner mask. Figure 5(b) reports endpoint-delay error because a model with high classification accuracy may still be operationally weak if it recognizes the endpoint too late. The proposed model reaches 94.6% endpoint accuracy and an average endpoint delay of 7.9 s. Since endpoint windows occupy a small portion of the stream, class-imbalance research [29] supports special attention to rare endpoint states, and explainability assessment work [30] supports reporting both delay and accuracy when model output affects process decisions. Transferability research [31] also helps explain why generic classification backbones may perform poorly in delayed endpoint settings.

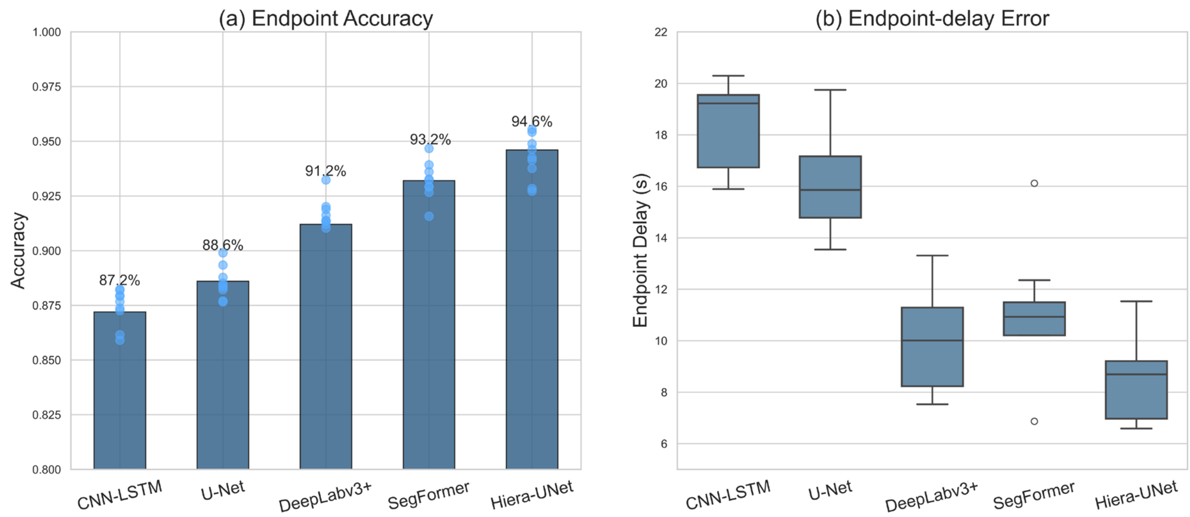

Figure 5. Endpoint-recognition performance. (a) Endpoint accuracy. (b) Endpoint-delay error.

Boundary-sensitive segmentation and temporal smoothing are both required for reliable endpoint recognition. SegFormer performs well on crisp images, but endpoint delay increases in highly blurred frames. DeepLabv3+ recovers large crystal regions, but it can lose fine nuclei during early growth transition. Explainable neural-network studies [32] support connecting model output with interpretable evidence, and contrastive

https://doi.org/10.64972/dea.2023.v2i3.3715d:52-64

representation research [33] supports using compact morphology differences to maintain stage separation. The proposed approach explains why endpoint probability increases: a successful prediction shows stable mask area, reduced boundary fluctuation, and high near-endpoint confidence. A visually good mask alone is not enough because the endpoint is a process state rather than a physical label printed on a single image.

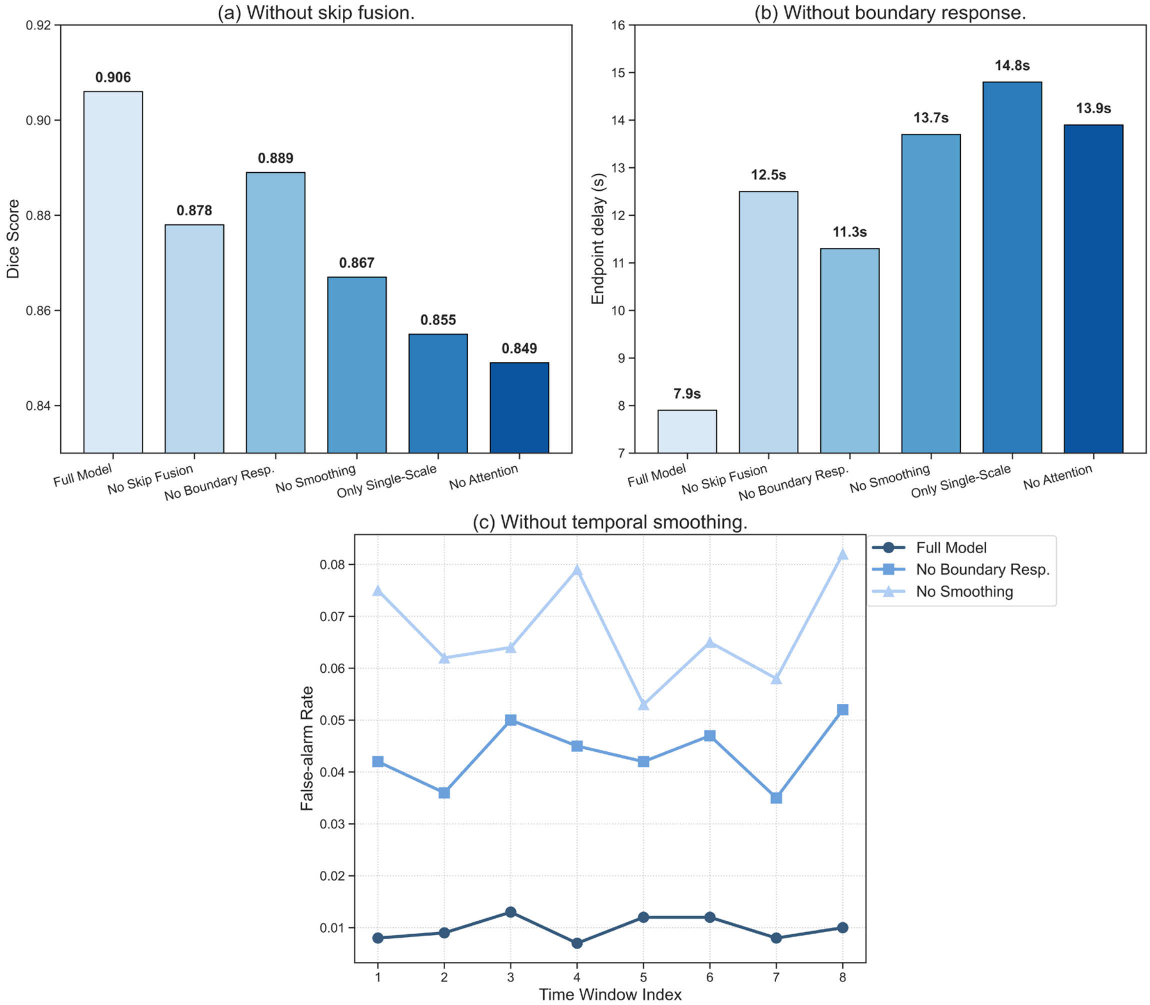

Figure 6. Ablation study of the endpoint-recognition model. (a) Without skip fusion. (b) Without boundary response. (c) Without temporal smoothing.

A few failed instances are also instructive. Sometimes, when there are noticeable bubbles near the probe window in the picture, U-Net misidentifies them as particles and widens the mask area. CNN-LSTM occasionally forecasts the ending too soon when the picture is extremely dense because the overall texture seems steady even if the crystal edges are still moving. Hiera-UNet concurrently uses temporal probability smoothing, boundary response, and mask shape to mitigate the errors. For near-saturation of the frame or abrupt agglomeration breakdown in the simulated batch, it is still not optimal. These examples suggest that in a real deployment, probe-state checks should be used in conjunction with the visual endpoint. In order to illustrate how the model functions in a process-control setting, failure analysis will also be covered in this section. The circular edges of bubble regions might seem as crystal clusters after reflection, which makes them challenging. The border branch loses high-frequency information in saturated frames, making them unsuitable for reliable masking. Another issue is abrupt agglomerate breakdown; once the majority of the batch has been processed, the picture may momentarily shift away from an endpoint-like structure. These examples show where a vision-only endpoint recogniser should be backed by probe-state information, illumination diagnostics, or a straightforward operator confirmation rule rather than invalidating the technique. For noisy but structurally consistent visual input, the model performs better. For images with ambiguous forms, it is not particularly

https://doi.org/10.64972/dea.2023.v2i3.3715d:52-64

effective. In order to prevent using the model as a "black box" for alerts, this division might be employed in deployment. Safer operator review is likewise supported by the same difference.

Ablation, Robustness, and Online Deployment Analysis

The ablation study is shown in Figure 6. Figure 6(a) removes hierarchical skip fusion to test whether small crystal edges can still be recovered without multi-level feature transfer; Dice drops from 0.906 to 0.878 because thin borders and fine crystals are missed. Figure 6(b) removes the boundary response, increasing endpoint delay by 4.6 s because the endpoint layer loses information about whether particle outlines have stabilized. Figure 6(c) removes temporal smoothing and produces several early false alarms during mixing disturbances, showing that a single-frame confidence value is not sufficient for online monitoring. Together, the three subfigures show that skip fusion preserves fine morphology, boundary response supports endpoint evidence, and smoothing converts image recognition into stable process monitoring.

Online robustness and deployment are displayed in Figure 7. Figure 7(a) reports endpoint accuracy under illumination noise and tests whether recognition depends too strongly on ideal imaging conditions. Under moderate exposure disturbance, the proposed model keeps accuracy above 91.0%, while the CNN-LSTM baseline remains below 86.0%. Figure 7(b) reports per-frame latency; Hiera-UNet maintains less than 22 ms on a desktop and less than 61 ms on an industrial edge device, which is necessary for online monitoring. Figure 7(c) shows an online warning curve connecting numerical endpoint probability with an operator-visible process signal. Reliable monitoring in disturbed operating environments has been emphasized in practical monitoring research [34], and broader AI deployment analysis [35] shows that technical models are more useful when they fit an actual decision process.

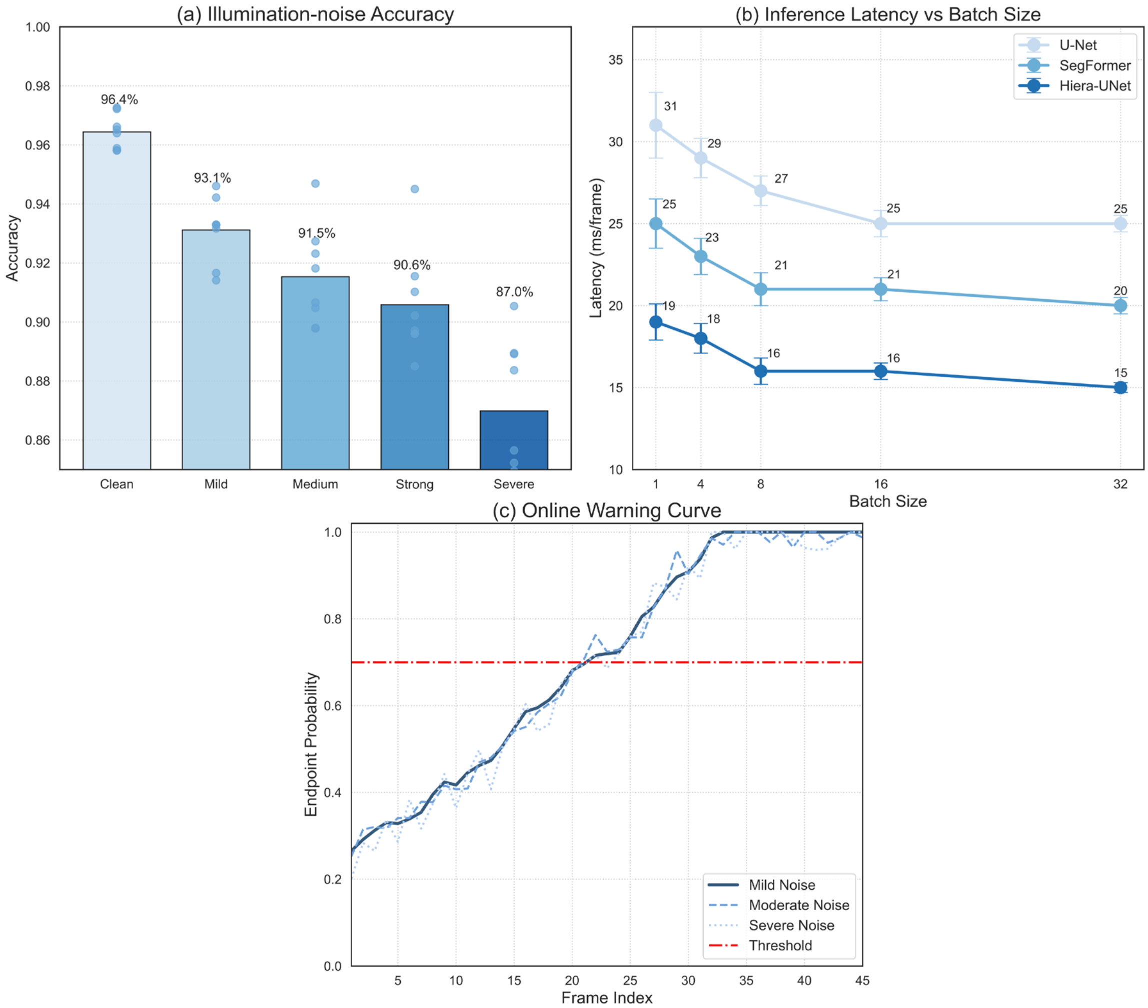

Figure 7. Online robustness and deployment interpretation. (a) Illumination-noise accuracy. (b) Inference latency. (c) Online warning curve.

https://doi.org/10.64972/dea.2023.v2i3.3715d:52-64

At the endpoint, the deployment value is most visible. Based on the image's categorisation findings, an operator must then determine whether to halt, hold, or continue the batch. Hiera-UNet produces a boundary-stability cue, a particle mask, and an endpoint probability curve. The outputs mentioned above will be merged. Examine the raw picture or probe state instead of making a choice if the endpoint probability is high but the mask contains bubble artefacts or saturated probe reflections. The endpoint decision gets more support if the probability steadily rises and the mask remains constant across successive windows. As a result, rather than acting as an independent batch-halt mechanism, the model will enhance decision-making.

What the model is not for has also been made clear. Do not take the place of laboratory confirmation or crystallisation expertise. For the end-point judgement, it will offer a uniform image-based helper. The operator will view the picture stream before taking action if the endpoint probability is high and the mask is still unstable. It could still be in a development stage with weak edge alterations if the mask remains steady and the likelihood is still low. Because they isolate segmentation evidence from the ultimate process decision, paired outputs of this kind are easier to utilise than single alerts. A generic process analytical equipment system for screening and alarm functions will be the practical application of the concept. Since the boundaries establish a safe-operation boundary, they are part of the technical contributions. As a result, without calibration for compound type, imaging geometry, and process recipe, it cannot be utilised as a general-purpose crystallisation controller. Additionally, it shouldn't be anticipated that it will resolve issues with crystals covered by chronic fouling or complete saturation of the probe picture. Serving as the visual endpoint for process recognition, connecting with process analytical technologies, past batch records, and operator judgement would be a more practical use. In this capacity, the model will extract evidence and eliminate the requirement for ongoing manual picture analysis. By preserving the period of morphological stabilisation and the decision's confidence, it can also carry out batch-to-batch comparisons. The programs are still extremely helpful for in-process control and end-of-line monitoring. This border also explains why discussion figures rather than a single final score are displayed in the study. Poor particle segmentation, an excessively sensitive reaction to a transient perturbation, a delayed reaction after reaching the real endpoint, or being too sluggish for the imaging interval are some of the reasons why a model in online crystallisation might fail. These failure scenarios cannot be distinguished by a single accuracy value. Each error's location and the model component that fixes it are displayed in figure-based analysis. Because of this, the suggested approach is more appropriately seen as a comprehensive monitoring plan than as a separate segmentation model. Image evidence, temporal continuity, end-point interpretation, and traceable batch supervision are all observable during operation.

Conclusion

In this study, a Hiera-UNet model based on online particle pictures is proposed for endpoint recognition in batch crystallisation. The technique reconstructs particle masks, creates normalised picture patches from the stream, extracts hierarchical crystal-morphology characteristics, and estimates endpoint probability via temporal smoothing. The Design connects segmentation evidence, boundary stability, morphological cues, and process-stage identification instead of seeing endpoint detection as a single-frame classification problem.

The simulated experiment shows that when compared to U-Net, UNet++, DeepLabv3+, SegFormer, and CNN-LSTM baselines, the suggested approach enhances mask Dice, endpoint accuracy, and endpoint-delay stability. Skip fusion, boundary response, and temporal smoothing all contribute to the ultimate outcome based on the ablation data. For online batch processing, this actually results in a stable warning curve and a clear particle mask.

There are still a few restrictions. Real online probe imaging from industrial crystallisation has not yet been included to the study, which uses simulated endpoint labels. Unusual light, probe fouling, bubbles, saturated reflections, and compound-specific crystal habits are all possible in a genuine system. Real-world online microscopy or probe-image streams, recipe-specific calibration, operator-in-the-loop endpoint decisions, and quality results from later batches will all be used in future work to verify the model. The suggested architecture is nevertheless a workable method of integrating online particle imaging, hierarchical segmentation, and endpoint identification for daily process supervision despite the drawbacks.

Author Contributions

Hrvoje Petković1 contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Ante Španjol contributes to analysis, investigation, data collection, draft preparation. and Domagoj Mihalić contributes to software, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2021). Deep learning for AI. Communications of the ACM, 64(7), 58-65. https://doi.org/10.1145/3448250

[2] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. Computational Intelligence and Neuroscience, 2018, 7068349. https://doi.org/10.1155/2018/7068349

[3] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., et al. (2020). Deep learning for generic object detection: A survey. International Journal of Computer Vision, 128, 261-318. https://doi.org/10.1007/s11263-019-01247-4

[4] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6, 60. https://doi.org/10.1186/s40537-019-0197-0

[5] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing, 152, 166-177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

[6] Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., et al. (2021). A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, 109(5), 756-795. https://doi.org/10.1109/JPROC.2021.3052449

[7] Arrieta, A. B., Diaz-Rodriguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable artificial intelligence: Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[8] Ilyas, I. F., & Rekatsinas, T. (2022). Machine learning and data cleaning: Which serves the other? Journal of Data and Information Quality, 14(3), Article 13. https://doi.org/10.1145/3514228

[9] Guo, M. H., Xu, T. X., Liu, J. J., Liu, Z. N., Jiang, P. T., Mu, T. J., et al. (2022). Attention mechanisms in computer vision: A survey. Computational Visual Media, 8, 331-368. https://doi.org/10.1007/s41095-022-0271-y

[10] Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. ACM Computing Surveys, 54(10s), Article 200. https://doi.org/10.1145/3505244

[11] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. European Conference on Computer Vision, 801-818. https://doi.org/10.1007/978-3-030-01234-2_49

[12] Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., de Lange, T., Johansen, D., & Johansen, H. D. (2020). DoubleU-Net: A deep convolutional neural network for medical image segmentation. 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems, 558-564. https://doi.org/10.1109/CBMS49503.2020.00111

[13] Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L., & Weinberger, K. Q. (2022). Convolutional networks with dense connectivity. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12), 8704-8716. https://doi.org/10.1109/TPAMI.2019.2918284

[14] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 11045, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1

[15] Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., et al. (2019). Attention gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis, 53, 197-207. https://doi.org/10.1016/j.media.2019.01.012

[16] Ibtehaz, N., & Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Networks, 121, 74-87. https://doi.org/10.1016/j.neunet.2019.08.025

[17] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11976-11986. https://doi.org/10.1109/CVPR52688.2022.01167

[18] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., et al. (2019). Searching for MobileNetV3. Proceedings of the IEEE/CVF International Conference on Computer Vision, 1314-1324. https://doi.org/10.1109/ICCV.2019.00140

[19] Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z., et al. (2021). Tokens-to-Token ViT: Training vision transformers from scratch on ImageNet. Proceedings of the IEEE/CVF International Conference on Computer Vision, 558-567. https://doi.org/10.1109/ICCV48922.2021.00060

[20] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, 9992-10002. https://doi.org/10.1109/ICCV48922.2021.00986

[21] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2021). The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. International Journal of Computer Vision, 129, 1038-1059. https://doi.org/10.1007/s11263-020-01400-4

[22] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. European Conference on Computer Vision, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13

[23] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2020). Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 318-327. https://doi.org/10.1109/TPAMI.2018.2858826

[24] Zhang, H., Sindagi, V., & Patel, V. M. (2020). Image de-raining using a conditional generative adversarial network. IEEE Transactions on Circuits and Systems for Video Technology, 30(11), 3943-3956. https://doi.org/10.1109/TCSVT.2019.2920407

[25] Li, B., Gou, Y., Gu, S., Liu, J. Z., Zhou, J. T., & Peng, X. (2021). You only look yourself: Unsupervised and untrained single image dehazing neural network. International Journal of Computer Vision, 129, 1754-1767. https://doi.org/10.1007/s11263-021-01431-5

[26] Wang, W., Xie, E., Li, X., Fan, D. P., Song, K., Liang, D., et al. (2021). Pyramid Vision Transformer: A versatile backbone for dense prediction without convolutions. Proceedings of the IEEE/CVF International Conference on Computer Vision, 548-558. https://doi.org/10.1109/ICCV48922.2021.00061

[27] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. European Conference on Computer Vision, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[28] Kirillov, A., He, K., Girshick, R., Rother, C., & Dollar, P. (2019). Panoptic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9404-9413. https://doi.org/10.1109/CVPR.2019.00963

[29] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6, 27. https://doi.org/10.1186/s40537-019-0192-5

[30] Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. Information Fusion, 76, 89-106. https://doi.org/10.1016/j.inffus.2021.05.009

[31] Kornblith, S., Shlens, J., & Le, Q. V. (2019). Do better ImageNet models transfer better? Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2661-2671. https://doi.org/10.1109/CVPR.2019.00272

[32] Le-Khac, P. H., Healy, G., & Smeaton, A. F. (2020). Contrastive representation learning: A framework and review. IEEE Access, 8, 193907-193934. https://doi.org/10.1109/ACCESS.2020.3031549

[33] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Muller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247-278. https://doi.org/10.1109/JPROC.2021.3060483

[34] Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8), 832. https://doi.org/10.3390/electronics8080832

https://doi.org/10.64972/dea.2023.v2i3.3715d:52-64

[35] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., et al. (2021). Artificial Intelligence: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002