Trypanosoma cruzi Detection using LSTM Convolutional Autoencoder

Geovani L. Martins[1], Daniel S. Ferreira[2], Claudia M. Carneiro[3], Andrea G. C. Bianchi[1,4]

[1] Postgraduate Program in Computer Science, Federal University of Ouro Preto, Ouro Preto, MG, Brazil

[2]Computing Department in Federal Institute of Education, Science, and Technology of Ceará, Maracanaú, CE, Brazil

[3]Clinical Analysis Department in Federal University of Ouro Preto, Ouro Preto, MG, Brazil

[4]Computing Department in Federal University of Ouro Preto, Ouro Preto, MG, Brazil

geovani.martins@aluno.ufop.edu.br, daniels@ifce.edu.br, {carneirocm,andrea}@ufop.edu.br

Abstract. *The presence of Trypanosoma cruzi (T. cruzi) parasites in blood samples is proof of the medical diagnosis of Chagas disease. Since the motion of these microorganisms is conspicuous in optical microscopy videos, we propose a spatio-temporal autoencoder for anomaly detection caused by parasite motility. This approach includes a spatial feature extractor and a temporal sequencer ConvLSTM for learning the temporal evolution of the spatial features. We trained the autoencoder with no parasites videos to learn the normal pattern and measured the regularity score in test videos with parasites. Our results showed that an LSTM-based autoencoder may identify T. cruzi anomalous motion, being a promising method for detecting parasites in microscopy videos.*

1. Introduction

Trypanosoma cruzi (*T. cruzi*) is the etiologic agent of Chagas disease [Pérez-Molina and Molina 2018]. The morphological and motile features of *T. cruzi* are key factors for identifying these microorganisms in laboratory inspections, such as in direct parasitological exams for medical diagnosis of Chagas disease, in which the microscopist evaluates the presence of parasites in peripheral blood samples. Once *T. cruzi* is visualized; there is no doubt about the positivity of the sample. However, some errors during the analysis can result in a wrong medical report, such as the incorrect handling of equipment, the inexperience of the microscopist, eye strain, and others.

The occurrence of false negatives in diagnosing Chagas disease has motivated some researchers to propose computational approaches for *T. cruzi* detection. These approaches can be divided into spatial [Nohara et al. 2010, Soberanis-Mukul et al. 2013, Uc-Cetina et al. 2013, Maza-Sastre et al. 2014, Moon et al. 2014, Uc-Cetina et al. 2015, Lakshmi et al. 2016, Relli et al. 2017, Ojeda-Pat et al. 2020, Vega-Alvarado et al. 2020, Pereira et al. 2019] and spatio-temporal [Alanis et al. 2004, Romero et al. 2012, Zhang et al. 2018, Takagi et al. 2019, Martins et al. 2021] according to the features investigated for the location of parasites. Approaches based on spatial features explore intrinsic aspects of microorganisms' color, texture, or morphology. Spatio-temporal

analysis approaches consider spatial information associated with temporal features in the scene, such as parasitic motion. Although spatio-temporal analysis for *T. cruzi* detection has been investigated recently, most of the works published in the last two decades report spatial features, as shown in Figure 1. In our literature review, only five computational approaches emphasize *T. cruzi* motion.

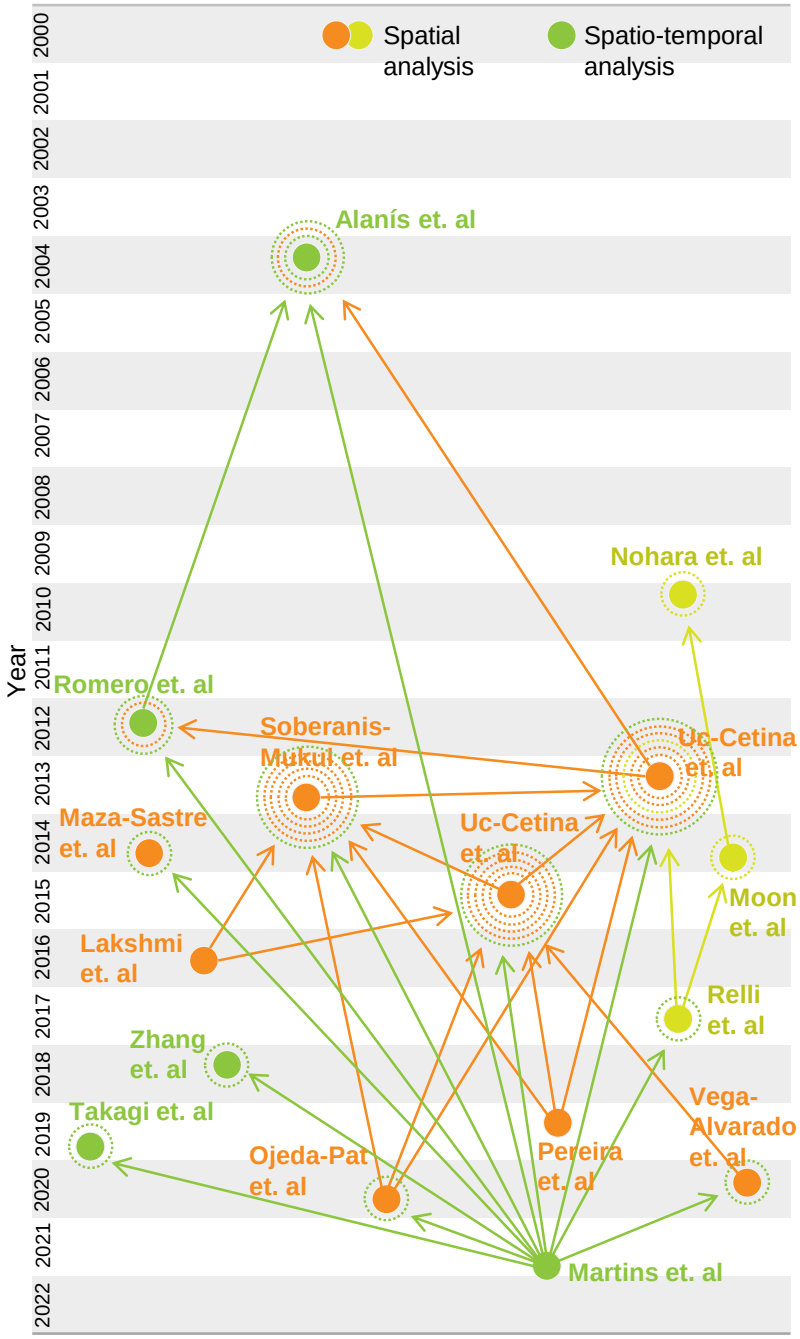

Figure 1. Historical direct citation network of *T. cruzi* detection works. Each node in the figure represents an article. The color indicates the work category. The arrows signal relationships (citation). The circles around the works indicate the number of citations that the article obtained. Note that there is a subdivision in the spatial analysis category (orange = extracellular detection; yellow = intracellular detection).

One of the challenges in *T. cruzi* detection by traditional computational approaches is the low contrast of parasites in medical images. These microorganisms are transparent in blood samples, making detection based on spatial characteristics difficult. This limitation favors the proposition of applications focused on motion, taking advantage of this intrinsic characteristic of the parasite. In optical microscopy videos, its motion tends to be salient, enabling detection [Martins et al. 2021]. The parasite's interactions with the blood cells cause various distortions in the motion of the *T. cruzi*, increasing intensity and randomness. The blood cells, although inert, tend to move under the action of dynamic stimuli, such as fluid displacement on the slide or microscope focus adjustments. If the normal pattern of the scene is to observe only cells in the microscope's visual field, the presence of parasites can be considered an abnormal event. In this context, the apparent motion of *T. cruzi* is seen as anomalous.

It is not trivial to define patterns for anomalies, as they are unpredictable. Some authors consider the detection of anomalies as a binary classification problem (normal and abnormal), which requires an accurate description of events. On the other hand, some approaches have implemented unsupervised models, such as autoencoders [Sabokrou et al. 2015, Hasan et al. 2016, Chong and Tay 2017, Wang et al. 2018]. Autoencoders are artificial neural networks that are trained to reconstruct the input. The idea is to design an architecture that has a code that can represent the original signal. These methods have two parts: the encoder learns representations of the input and sends the learned signal to the code layer; the decoder produces the signal reconstruction from the code layer. Autoencoders can learn spatial and spatio-temporal information from unlabeled video clips of normal scenes or with few anomalies.

We hypothesize that autoencoders can be used to detect *T. cruzi* parasites in optical microscopy videos, considering that the motion of these microorganisms is anomalous. To confirm this hypothesis, we propose a convolutional spatio-temporal autoencoder to learn the regular dynamic pattern in training videos and then to be able to detect when an abnormal event occurs in test videos. To the best of our knowledge, the use of autoencoders to detect *T. cruzi* based on the identification of abnormal motion patterns is novel. Our main contribution lies in detect parasites based on an approach focused on anomaly detection according to methodoly described in Section 2. The experimental results are discussed in Section 3, and Section 4 presents our conclusions.

2. Methods

Our approach using a convolutional long short-term memory (ConvLSTM) autoencoder is shown in Figure 2. We detail the proposed approach below.

2.1. Preprocessing

We extracted video clippings by [Martins et al. 2021], generating a new dataset for this work. The video clippings have dimensions of 80 x 80 pixels, and the regions were chosen according to the step of the experiment. In training, we obtained 30 video clippings, in which the regions were randomly selected, excluding those with parasites. Therefore, only video clippings with a regular dynamic pattern. Each training video clip has 200 frames, resulting in a dataset with 6000 frames. The test video clips alternate between regular scenes and scenes with abnormal events, that is, with anomalous motions of the

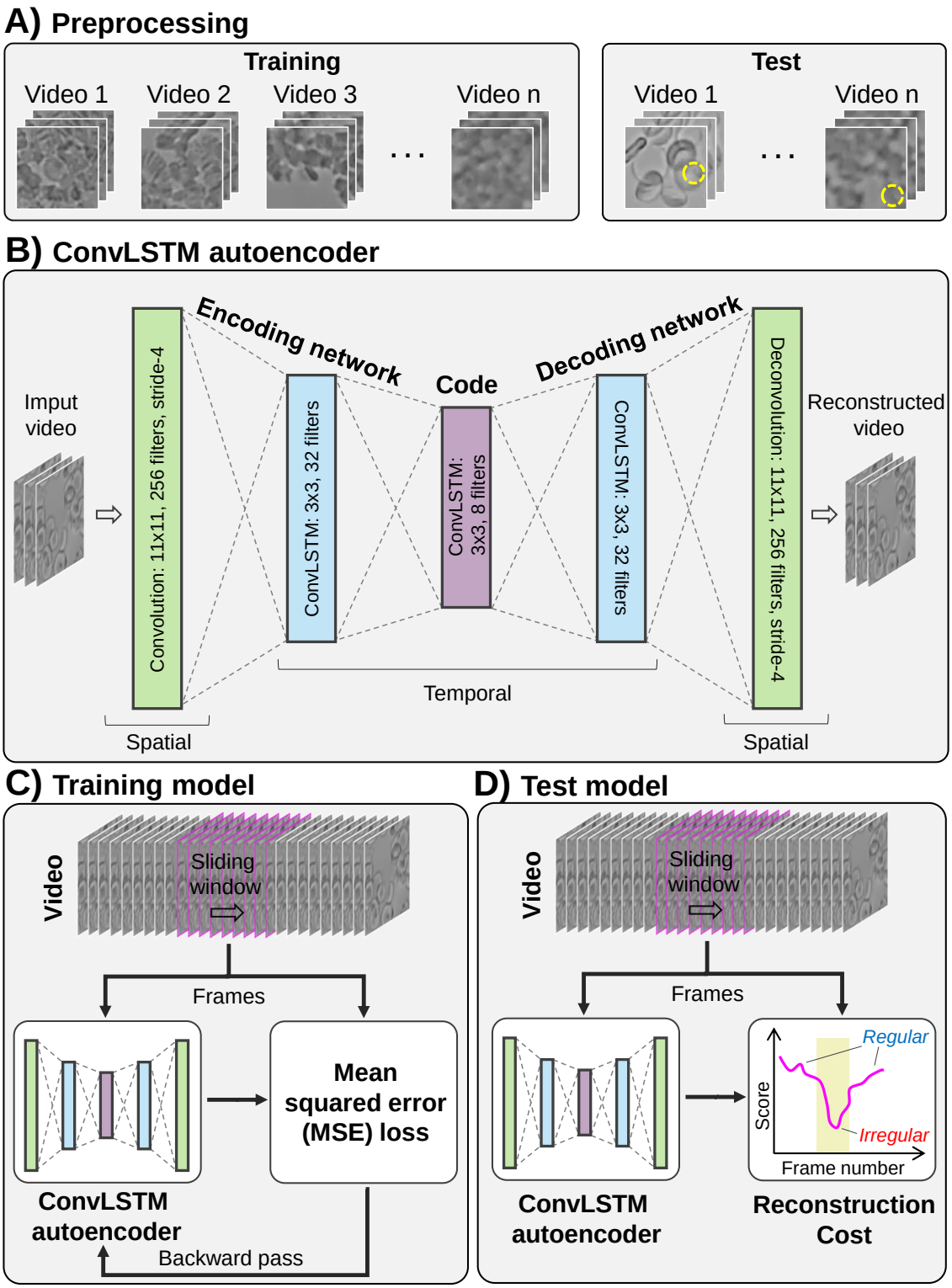

Figure 2. Overview of our approach. (A) Pre-processing of the optical microscopy videos. (B) Convolutional long short-term memory (convLSTM)-based autoencoder architecture. (C) Model training process. (D) Model test process.

parasites. We obtained 5 video clippings without limiting the total frames, generating 1169 test video frames. To normalize the frame videos, each pixel value is divided by

255, to obtain a value between [0; 1]. After that, the frames are converted to grayscale to reduce dimensionality.

2.2. ConvLSTM autoencoder

We introduce an autoencoder based on deep convolutional long short-term memory (ConvLSTM) to learn regularity in optical microscopy videos. It consists of two parts: spatial autoencoder for learning spatial information (e.g., color, texture, shape) and temporal encoder-decoder for learning temporal information (e.g., cell and parasite motions, the dynamism of microscope focus adjustments).

The proposed ConvLSTM autoencoder has an architecture inspired by [Chong and Tay 2017]. However, our proposal has fewer layers and parameters. Different from [Chong and Tay 2017], we employ a single convolutional and deconvolutional layers to form the spatial encoder and decoder, respectively. The convolutional layer computes the convolutional operation of the input frames using kernel filters to extract fundamental features, preserving the spatial relationship between the pixels and downsampling the input. A smaller number of filters is desired, avoiding a large computational effort. The deconvolution layer performs a mathematical operation that reverses the process of a convolutional layer [Zeiler et al. 2010]. We use hyperbolic tangent as the activation function of spatial-temporal encoder and decoder. A final convolutional layer generates the output. It consists of 1 filter and sigmoid activation function with kernel size (11, 11).

We propose three convolutional LSTM layers to solve the temporal dependency in the data. The ConvLSTM extends the fully connected LSTM (FC-LSTM), combining convolution and recurrent LSTM networks. It replaces the matrix multiplication operations used in LSTM with convolution operations [Shi et al. 2015]. ConvLSTM requires fewer weights and yields better spatial feature maps. Except for the code layer, we use a dropout rate of 20% in the ConvLSTM layers. It is a technique to reduce overfitting [Srivastava et al. 2014], also indicated for training a deep network with a relatively small dataset. We apply layer normalization after each convolution or ConvLSTM. In recurrent networks, layer normalization performs better than batch normalization [Ba et al. 2016].

2.3. Autoencoder workflow

The autoencoder workflow involves training and testing. The optical microscopy video frames are divided into 20-sized temporal sequences using the sliding window technique. We perform data augmentation in the temporal dimension for training, generating more sequences. We concatenate frames with various skipping strides, as proposed by [Hasan et al. 2016]. The model is trained by minimizing the reconstruction error of the input volume. The reconstruction error is expressed as the mean square error (MSE). Adam optimization [Kingma and Ba 2015, Hassan et al. 2023] is implemented to provide more efficient neural network weights by running repeated cycles of adaptive moment estimation. We adjust the layer weights based on the loss function, applying backpropagation and updating the weights in the multilayers following the MSE reduction. Once the model is trained, we can test each testing video individually. We use the sliding window technique to get all the consecutive 20-frame sequences. After learning the spatio-temporal regularity in optical microscopy videos, our autoencoder can reconstruct video sequences without parasites with low error, as opposed to video sequences with the presence of moving parasites.

2.4. Abnormality evaluation

2.4.1. Mean squared error (MSE)

We use the MSE as a loss function. Loss functions are a measurement of how good our model is in terms of predicting the expected outcome. The MSE is calculated as the average of the squared differences between the predicted $(\hat{y})$ and actual values (y), as shown in Equation 1:

$$MSE(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} (y - \hat{y}_i)^2, \tag{1}$$

where N is the number of samples we are training against. The MSE is always positive, though it can be 0 if the predictions are completely accurate. Notice that the MSE loss function penalizes the model for making large errors by squaring them.

2.4.2. Regularity Score

The reconstruction error of a pixel's intensity value I at the location (x, y) in a given video frame t is obtained using the L2 norm as shown in equation 2.

$$e(x, y, t) = ||I(x, y, t) - f_w(I(x, y, t))||_2, \tag{2}$$

where f_w is the learned model by the ConvLSTM autoencoder. Then, we compute the reconstruction error of a frame t by summing up all the pixel-wise errors:

$$e(t) = \sum_{(x,y)} e(x, y, t). \tag{3}$$

We estimate the abnormality score $s_a(t)$ by scaling between 0 and 1:

$$s_a(t) = \frac{e(t) - e(t)_{min}}{e(t)_{max}}. \tag{4}$$

Finally, we can derive regularity score $s_r(t)$ by subtracting abnormality scores from 1:

$$s_r(t) = 1 - s_a(t) \tag{5}$$

Video sequences containing normal events have a higher regularity score since they are similar to the data used to train the model. In comparison, sequences containing abnormal events have a lower regularity score.

3. Results and Discussion

We compare our results with Chong's work [Chong and Tay 2017], which is considered a state-of-the-art approach to video anomaly detection. The results are presented and discussed below.

3.1. Model loss

Figure 3 shows the cost function of the models we trained. We implement Adam optimizer with a learning rate set to 10^{-4}, we reduce it when training loss stops decreasing by using a decay of 10^{-5}, and we set the epsilon value to 10^{-6}. We use mini-batches of size 8 and each training volume is trained for a maximum of 20 epochs. Note that both models converge quickly, before 20 epochs.

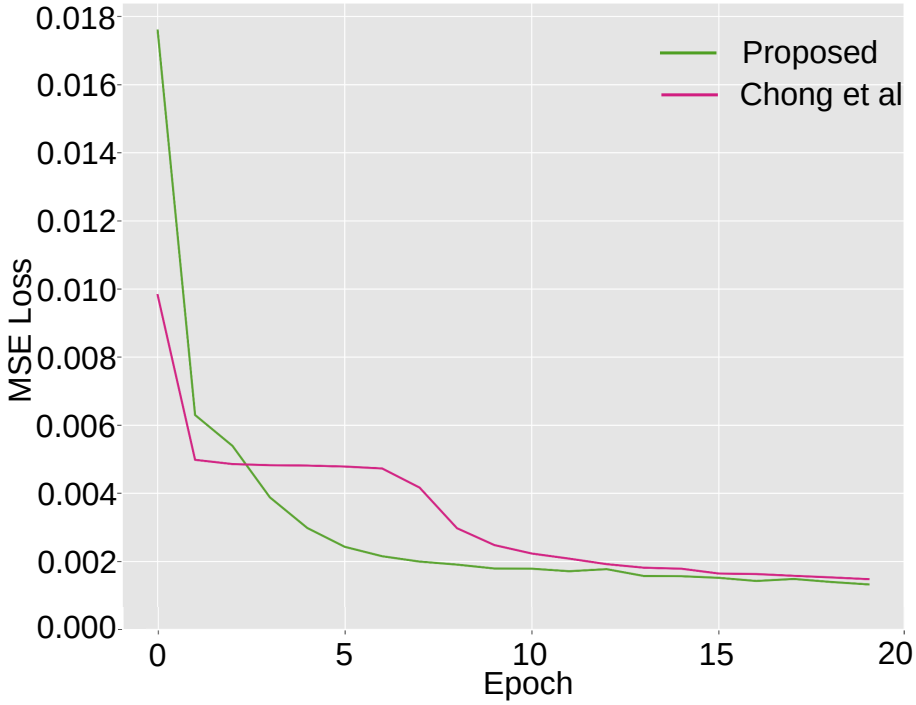

Figure 3. Evolution of model's loss (training).

3.2. Scene regularity

A comparison of predictions between normal and abnormal events is shown in Figure 4. As Figure 4A shows, our method can produce higher regularity scores during normal activities and lower scores when abnormalities occur. We observe the parasite completely in the scene, and its interactions with blood cells allow its detection. Our approach outperforms Chong's method in the presence of the *T. cruzi* since the regularity score was stable in a range of lower values.

Figure 4B shows another scenario in optical microscopy videos. The *T. cruzi* parasite does not appear completely in the scene, being observed at the bottom of the video. Our approach detected the abnormality and performed better than Chong's method. In the normal section of the video, we also stabilized the regularity score in a range of higher values. Note that there is a focus adjustment during the transition from normal to abnormal scenes, resulting in the blurring of the video frame. Both methods have difficulty keeping the regularity score at low values in this scenario. In addition, the parasite not being fully visible in the scene also contributed to this result. In this case, interactions with cells at the edge of the video are crucial for detection.

The results were similar in Figure 4C. Both approaches obtain a peak of the regularity score in the section without parasites. *T. cruzi* interacts with some cells at the edge

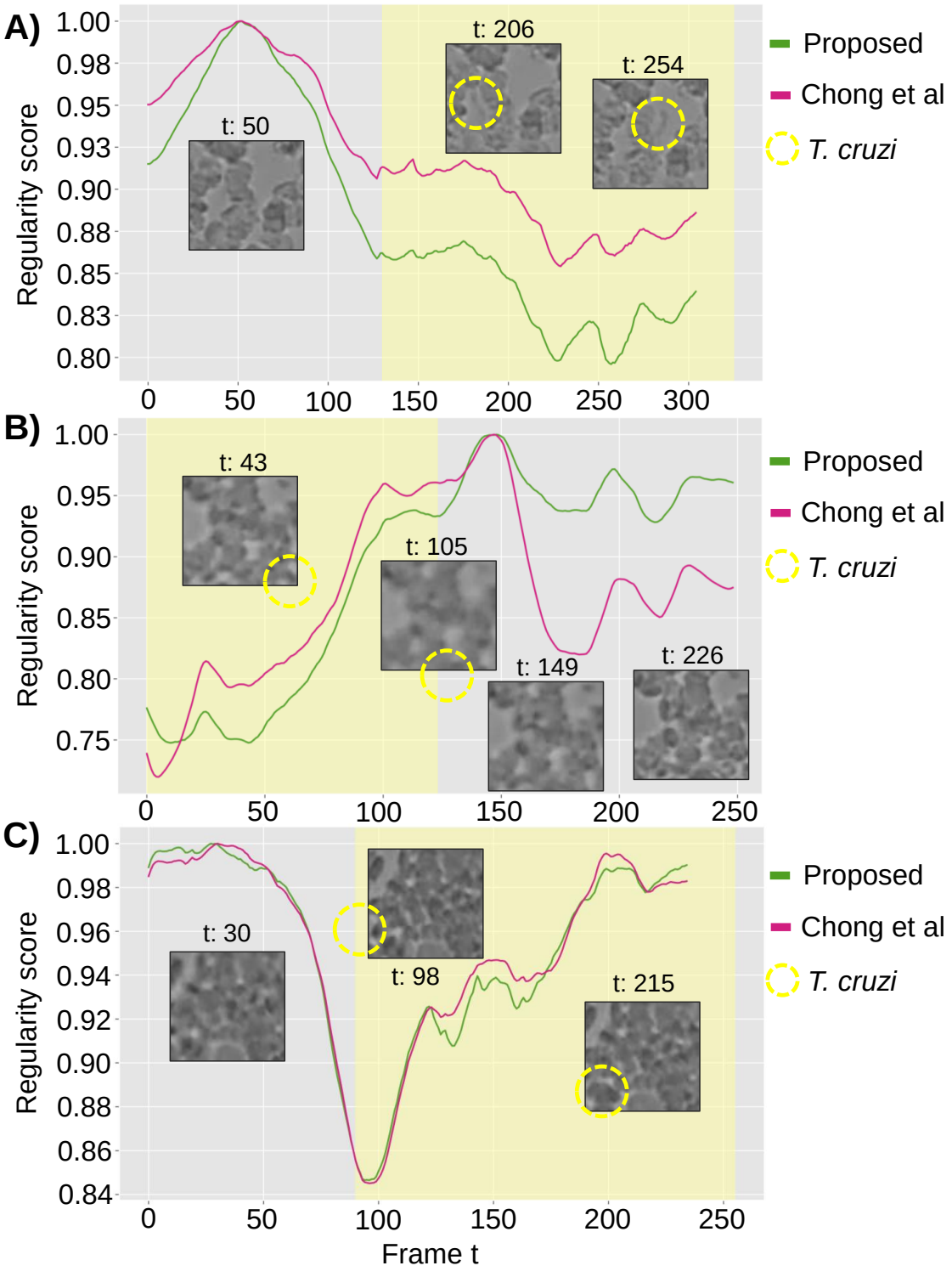

Figure 4. Regularity score. A) - C) Optical microscopy videos. Yellow sections indicate the presence of *T. cruzi* parasites.

of the video. After that, it appears to interact with a cluster of cells. Both approaches incorrectly indicate normality at t = 200. We observed that the video has high blurring, contributing to this result.

In Figure 5, we analyze the results of detecting abnormalities in test videos with spatial characteristics slightly different from those used for training. The frames of these videos show more sharper contour blood cells. Figure 5A shows that our approach obtained a higher regularity score than Chong's method in the normal section of the video. Then, the two approaches showed similar behavior in the abnormal section. The peak of the regularity score in the abnormal section is a false negative. *T. cruzi*'s interactions with the neighboring cell result in parasite occlusion in some video frames, as observed in frame 100. Note that the parasite's cell body overlaps the cell's contour, which may look like a normal scene. After that, we obtain a minimum regularity score. As expected in Figure 5B, the regularity score is minimal in the abnormal section of the video. However, we observed a curve oscillation between frames 35 and 70, incorrectly indicating an abnormality.

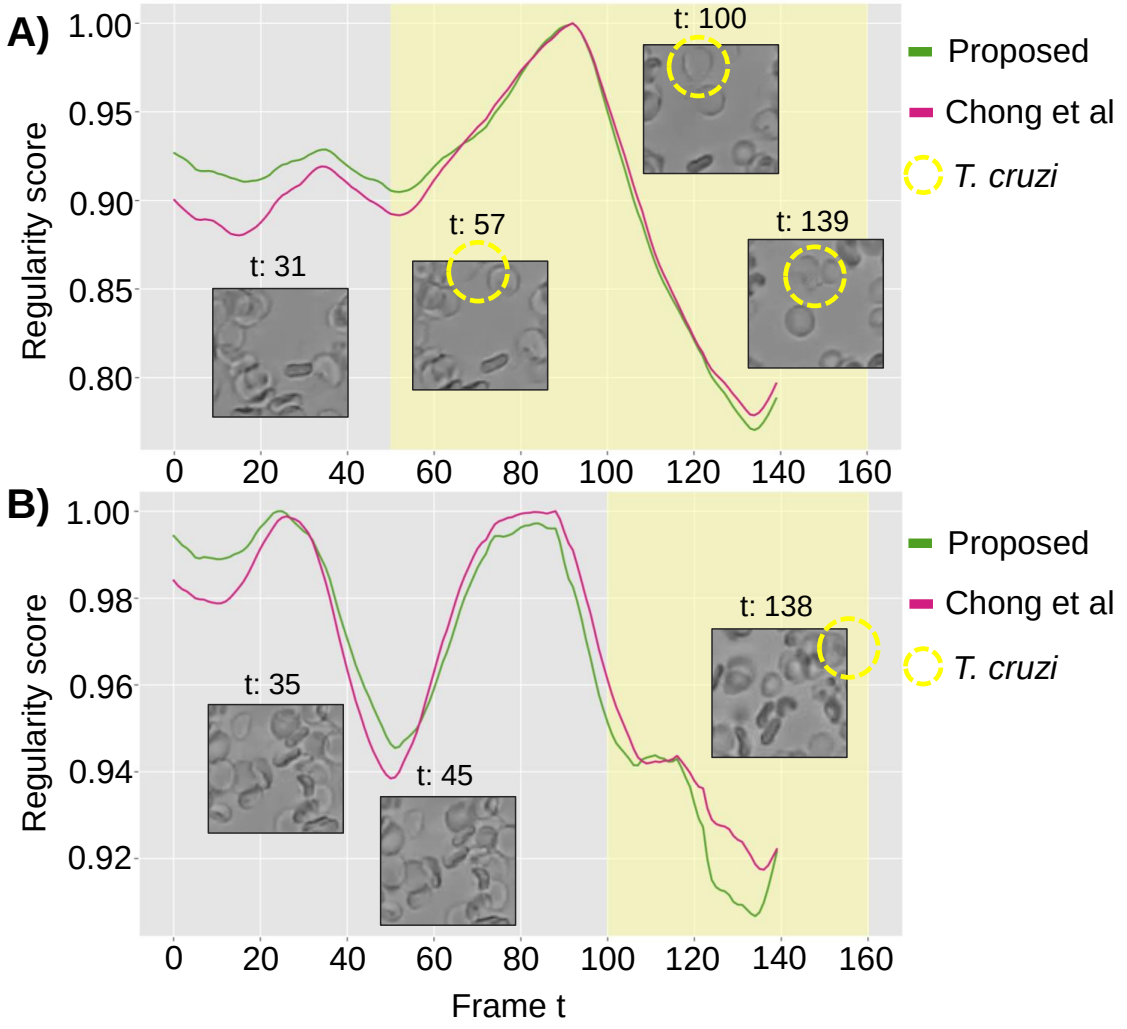

Figure 5. Regularity score in videos with distinct spatial characteristics. A) and B) Optical microscopy videos. Yellow sections indicate the presence of *T. cruzi* parasites.

Our results are promising since the images used are extremely complex from a computational point of view. The videos are obtained from *T. cruzi* parasitological analyzes using conventional optical microscopes. In addition to the low quality of the images,

the parasite has low contrast in the scene, and its motion is the main aspect that guides detection. It is important to emphasize that blood cells also have dynamism, contributing to false predictions. Predicting anomalous events in which the parasite's cell body resembles the cells' contour or during the occurrence of microscopic focus adjustments is quite challenging. The approaches consider the scene as normal, increasing the regularity score. Furthermore, parasites that do not entirely appear in the scene and interact with cells at the edge of the video need these cells to be stimulated sufficiently to indicate the anomaly. A limitation of our work is that we were unable to specify a threshold for the regularity score, aiming at a detection application. Future work with more data could help us in this direction.

4. Conclusion

We introduce an autoencoder based on LSTM convolutional (ConvLSTM) for *T. cruzi* detection in optical microscopy videos. Our approach focuses on motion since this microorganism has low contrast in these images. Since *T. cruzi* motion is salient in parasitological analysis, we consider the motion of this microorganism to be anomalous compared to the regular dynamic pattern of blood microscopy. Thus, our model is trained with videos of normal dynamic patterns of blood microscopy to minimize the reconstruction error between the input video and the output video reconstructed by the model. Then, we tested the model with videos that alternate between scenes with normal patterns (without parasites) and abnormal patterns (with parasites) so that the presence of *T. cruzi* incurs a low regularity score in the analyzed scene. The results showed that our model performs competitively with a state-of-the-art anomaly detection method, surpassing in some anomaly scenarios. These results are promising, since our approach has an architecture with fewer layers and parameters. For future work, some adjustments in the architecture of the neural network should be considered, in addition to an analysis with more video samples to specify a threshold for the regularity score in *T. cruzi* detection.

5. Acknowledgements

The authors would like to thank the Foundation for Research Support of the State of Minas Gerais (FAPEMIG) under grants 5.43/21, APQ-00751-19 and APQ-01306-22, the National Council of Scientific and Technological Development (CNPq) under grants 305895/2019-2, and the Dean of Research, Graduate Studies and Innovation (PROPPI/UFOP) under grants 23109.000928/2020-33 and 23109.000929/2020-88 for their financial support of this research.

References

Alanis, E., Romero, G., Alvarez, L., Martinez, C. C., and Basombrio, M. A. (2004). Optical detection of *Trypanosoma cruzi* in blood samples for diagnosis purpose. In *5th Iberoamerican Meeting on Optics and 8th Latin American Meeting on Optics, Lasers, and Their Applications*, volume 5622, pages 24–28, Porlamar, Venezuela. SPIE.

Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. arXiv:1607.06450.

Chong, Y. S. and Tay, Y. H. (2017). Abnormal event detection in videos using spatiotemporal autoencoder. In Cong, F., Leung, A., and Wei, Q., editors, *Advances in Neural Networks - ISNN 2017*, volume 10262, pages 189–196, Cham. Springer International Publishing.

Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A. K., and Davis, L. S. (2016). Learning temporal regularity in video sequences. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 733–742, Los Alamitos, CA, USA. IEEE Computer Society.

Hassan, E., Shams, M., Hikal, N., and Elmougy, S. (2023). The effect of choosing optimizer algorithms to improve computer vision tasks: A comparative study. *Multimed. Tools. Appl.*, 82:16591–16633.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA.

Lakshmi, V. S., Tebby, S. G., Shriranjani, D., and Rajinikanth, V. (2016). Chaotic cuckoo search and kapur/tsallis approach in segmentation of *T. cruzi* from blood smear images. *Int. J. Comp. Sci. Infor. Sec.*, 14:51–56.

Martins, G. L., Ferreira, D. S., and Ramalho, G. L. B. (2021). Collateral motion saliency-based model for *Trypanosoma cruzi* detection in dye-free blood microscopy. *Comput. Biol. Med.*, 132:104220.

Maza-Sastre, H., Ochoa-Montiel, R., Sánchez-López, C., Pérez-Corona, C., Carrasco-Aguilar, M. A., and Morales-López, F. E. (2014). Identification of trypanosoma with digital image processing. In *Proceedings of the 2014 IEEE Central America and Panama Convention (CONCAPAN XXXIV)*, pages 1–4, Panama City, Panama. IEEE.

Moon, S., Siqueira-Neto, J. L., Moraes, C. B., Yang, G., Kang, M., Freitas-Junior, L. H., and Hansen, M. A. E. (2014). An image-based algorithm for precise and accurate high throughput assessment of drug activity against the human parasite *Trypanosoma cruzi*. *PLoS One*, 9(2):e87188.

Nohara, L. L., Lema, C., Bader, J. O., Aguilera, R. J., and Almeida, I. C. (2010). High-content imaging for automated determination of host-cell infection rate by the intracellular parasite *Trypanosoma cruzi*. *Parasitol. Int.*, 59(4):565–570.

Ojeda-Pat, A., Martin-Gonzalez, A., and Soberanis-Mukul, R. (2020). Convolutional neural network U-Net for *Trypanosoma cruzi* segmentation. In *Intelligent Computing Systems*, volume 1187, pages 118–131, Cham. Springer.

Pereira, A., Pyrrho, A., Vanzan, D., Mazza, L., and Gomes, J. G. (2019). Deep convolutional neural network applied to chagas disease parasitemia assessment. In *Anais do 14 Congresso Brasileiro de Inteligência Computacional*, pages 1–8, Curitiba, Brazil. ABRICOM.

Pérez-Molina, J. A. and Molina, I. (2018). Chagas disease. *The Lancet*, 391(10115):82–94.

Relli, C. D. S., Facon, J., Ayala, H. L., and Britto, A. D. S. (2017). Automatic counting of trypanosomatid amastigotes in infected human cells. *Comput. Biol. Med.*, 89:222–235.

Romero, G. G., Monaldi, A. C., and Alanís, E. E. (2012). Digital holographic microscopy for detection of *Trypanosoma cruzi* parasites in fresh blood mounts. *Opt. Commun.*, 285(6):1613–1618.

Sabokrou, M., Fathy, M., Hoseini, M., and Klette, R. (2015). Real-time anomaly detection and localization in crowded scenes. In *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 56–62, Los Alamitos, CA, USA. IEEE Computer Society.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, page 802–810, Cambridge, MA, USA. MIT Press.

Soberanis-Mukul, R., Uc-Cetina, V., Brito-Loeza, C., and Ruiz-Piña, H. (2013). An automatic algorithm for the detection of *Trypanosoma cruzi* parasites in blood sample images. *Comput. Meth. Prog Bio.*, 112(3):633–639.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958.

Takagi, Y., Nosato, H., Doi, M., Furukawa, K., and Sakanashi, H. (2019). Development of a motion-based cell-counting system for trypanosoma parasite using a pattern recognition approach. *Biotechniques.*, 66(4):179–185.

Uc-Cetina, V., Brito-Loeza, C., and Ruiz-Piña, H. (2013). Chagas parasites detection through gaussian discriminant analysis. *Abstr. Appl.*, 8:6–17.

Uc-Cetina, V., Brito-Loeza, C., and Ruiz-Piña, H. (2015). Chagas parasite detection in blood images using adaboost. *Comput. Math. Methods Med.*, 2015:1–13.

Vega-Alvarado, L., Caballero-Ruiz, A., Ruiz-Huerta, L., Heredia-López, F., and Ruiz-Piña, H. (2020). Images analysis method for the detection of Chagas parasite in blood image. In *Pattern Recognition Techniques Applied to Biomedical Problems*, pages 63–72, Cham. Springer.

Wang, L., Zhou, F., Li, Z., Zuo, W., and Tan, H. (2018). Abnormal event detection in videos using hybrid spatio-temporal autoencoder. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2276–2280.

Zeiler, M. D., Krishnan, D., Taylor, G. W., and Fergus, R. (2010). Deconvolutional networks. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2528–2535.

Zhang, Y., Koydemir, H. C., Shimogawa, M. M., Yalcin, S., Guziak, A., Liu, T., Oguz, I., Huang, Y., Bai, B., Luo, Y., et al. (2018). Motility-based label-free detection of parasites in bodily fluids using holographic speckle analysis and deep learning. *Light Sci. Appl.*, 7(1):108.