IDEAPUZZLE
DESIGN AND DEFEND YOUR PHD

TML-deepLearning-LVP

Author: Lara Visuña
Date: 2023-06-22 - 09:34:49 (WET)
Self-evaluation: 86%

Key words

1/21. Theoretical question: what are the two main keywords of your research?

My thesis focuses on two main keywords "Deep learning" and "Medical images", this particular research focuses on "Timelapse Microscopy" (TML) images.
-Deep learning: It is a specific area of Artificial intelligence.
-TML: Timelapse microscopy is a technique of image acquisition with a microscope to periodically capture images to compose a video.
The title of my research design is "Deep Learning for the automatic bacilli monitoring in time-lapse microscopy (TLM) images".
Use case: TLM is used to evaluate bacterial culture evolution and response in different conditions, like applying different drugs [1]. My research focuses on speeding up this process to develop tuberculosis (TB) drugs, providing automatic monitoring for mycobacterial tuberculosis growth.

[1] Aknoun, S., Yonnet, M., Djabari, Z., Graslin, F., Taylor, M., Pourcher, T., ... & Pognonec, P. (2021). Quantitative phase microscopy for non-invasive live cell population monitoring. Scientific Reports, 11(1), 4409.
Self-evaluation: 100%

Self-evaluation: 100%

Streams of thought

2/21. Theoretical question: what are the two main streams of thought of your literature review?

Nowadays, the evaluation of TLM images is a manual process. To improve the sensibility and speed up this process there are many previous works, showing two main strengths.
On one hand, a big amount of work does individual cell tracking and evaluation, based on the singularity of every individual [1,2,3].
The other stream is based on the evaluation of the whole colony [4], there are studies that consider the global cell population an even more robust and accurate method to assess its growth parameters [5].
This work wants to mix the two streams to offer global and individual parameters of growth.

[1] Lugagne, J. B., Lin, H., & Dunlop, M. J. (2020). DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. PLoS computational biology, 16(4), e1007673.
[2] Tsai, H. F., Gajda, J., Sloan, T. F., Rares, A., & Shen, A. Q. (2019). Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning. SoftwareX, 9, 230-237.
[3] Ulicna, K., Vallardi, G., Charras, G., & Lowe, A. R. (2021). Automated deep lineage tree analysis using a Bayesian single cell tracking approach. Frontiers in Computer Science, 3, 734559.
[4] Wang, H., Ceylan Koydemir, H., Qiu, Y., Bai, B., Zhang, Y., Jin, Y., ... & Ozcan, A. (2020). Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. Light: Science & Applications, 9(1), 118
[5] Aknoun, S., Yonnet, M., Djabari, Z., Graslin, F., Taylor, M., Pourcher, T., ... & Pognonec, P. (2021). Quantitative phase microscopy for non-invasive live cell population monitoring. Scientific Reports, 11(1), 4409.

Self-evaluation: 100%

Research gap

3/21. Theoretical question: what is the main gap that your research addresses?

Deep learning has shown a huge amount of development, but there are still challenges remaining, especially in the biomedical area.
The main drawback in the TML area is the huge amount of annotated dataset needed, which requires tedious work slowing the development of new techniques [1]. We can find several papers claiming the development of deep learning in drug discovery and also microscopy images:
"Deep learning has achieved satisfactory results in classification and segmentation tasks. However, for particle tracking in microscopy images, especially for densely decorated multiple-target tracking, the performance of deep learning is still not sufficient. "[1]
"We hope that increased availability of high quality query able image datasets paired with side information on imaged compounds, genetic perturbations, or disease models will in turn inspire the design of yet more powerful machine-learning methods, driving a virtuous circle of discovery."[2]

[1] Liu, Z., Jin, L., Chen, J., Fang, Q., Ablameyko, S., Yin, Z., & Xu, Y. (2021). A survey on applications of deep learning in microscopy image analysis. Computers in Biology and Medicine, 134, 104523.
[2] Chandrasekaran, S. N., Ceulemans, H., Boyd, J. D., & Carpenter, A. E. (2021). Image-based profiling for drug discovery: due for a machine-learning upgrade?. Nature Reviews Drug Discovery, 20(2), 145-159.

Self-evaluation: 100%

Research question or hypothesis

4/21. Theoretical question: what is the main question or hypothesis of your research?

This research addresses the following broad research question: HOW deep learning can influence time-lapse microscopy images for speeding up bacterial analysis?

Self-evaluation: 50%

State of the science

5/21. Theoretical question: what is the current answer to your research question or hypothesis?

Drug discovery is beginning to be accelerated by deep learning tools [1]. Several efforts to detect bacillus and other cells in an automatic way with deep learning are currently reported in the available literature [2-8].
In [2] the UNet and the UNet++ architectures are used for the detection and segmentation of bacteria (B. anthracis) in microscopy images belonging to tissues of patients infected. Two UNet architectures are also used in [3], for the segmentation and tracking of single cells in TML images. An updated version of the software was presented in [4] for tracking and segmentation in 2D.
Authors in [5] develop a stain-free, single-cell segmentation and tracking , a semi-automatic system for phase contrast TLM images. A fully automatic method to track the cells over time is presented by Uclina et al [6] using phase contrast and fluorescence TLM sequences.

A colony-level detection algorithm is presented in [9], the deep learning algorithm computes and detects the level of growth of different bacterial species.

Despite the advantages of artificial intelligence, there are some previous works with remarkable results in cell detection thanks to processing techniques. Piersma et al. [10] apply image processing techniques to TMLs of bacillus subtilis to measure gene expression heterogeneity, using phase contrast and fluorescence channels in the culture images. In [11] cell detection and segmentation in time-lapse sequences are presented, with fully automatic processing image techniques, evaluating the results in 7 different TML datasets.

Existing literature presents several algorithms and methods available for TML bacilli tracking and segmentation, however, this research line continues to be a challenge. This research focuses on the design of a fully automated TML deep learning algorithm for the characterization of mycobacterium tuberculosis, which causes de tuberculosis disease.

[1] Chandrasekaran, S. N., Ceulemans, H., Boyd, J. D., & Carpenter, A. E. (2021). Image-based profiling for drug discovery: due for a machine-learning upgrade?. Nature Reviews Drug Discovery, 20(2), 145-159.
[2] Hoorali, F., Khosravi, H., & Moradi, B. (2020). Automatic Bacillus anthracis bacteria detection and segmentation in microscopic images using UNet++. Journal of Microbiological Methods, 177, 106056.
[3] Lugagne, J. B., Lin, H., & Dunlop, M. J. (2020). DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. PLoS computational biology, 16(4), e1007673.
[4] O?Connor, O. M., Alnahhas, R. N., Lugagne, J. B., & Dunlop, M. J. (2022). DeLTA 2.0: A deep learning pipeline for quantifying single-cell spatial and temporal dynamics. PLOS Computational Biology, 18(1), e1009797.
[5] Tsai, H. F., Gajda, J., Sloan, T. F., Rares, A., & Shen, A. Q. (2019). Usiigaci: Instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning. SoftwareX, 9, 230-237.
[6] Ulicna, K., Vallardi, G., Charras, G., & Lowe, A. R. (2021). Automated deep lineage tree analysis using a Bayesian single cell tracking approach. Frontiers in Computer Science, 3, 734559.
[7] Du, F., He, L., Lu, X., Li, Y. Q., & Yuan, Y. (2023). Accurate identification of living Bacillus spores using laser tweezers Raman spectroscopy and deep learning. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 289, 122216.
[8] Lo, C. M., Wu, Y. H., Li, Y. C., & Lee, C. C. (2020). Computer-aided bacillus detection in whole-slide pathological images using a deep convolutional neural network. Applied Sciences, 10(12), 4059.
[9] Wang, H., Ceylan Koydemir, H., Qiu, Y., Bai, B., Zhang, Y., Jin, Y., ... & Ozcan, A. (2020). Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. Light: Science & Applications, 9(1), 118.
[10] Piersma, S., Denham, E. L., Drulhe, S., Tonk, R. H., Schwikowski, B., & van Dijl, J. M. (2013). TLM-Quant: an open-source pipeline for visualization and quantification of gene expression heterogeneity in growing microbial cells. PLoS One, 8(7), e68696.
[11] Makrogiannis, S., Annasamudram, N., Wang, Y., Miranda, H., & Zheng, K. (2021, December). A system for spatio-temporal cell detection and segmentation in time-lapse microscopy. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 2266-2273). IEEE.

Self-evaluation: 50%

Philosophical stance

6/21. Methodological question: what is the philosophical stance of your research?

The philosophical stance of this research is quantitative objectivism. The research is focused on developing a deep-learning network to evaluate TLM images to produce an objective measure of bacterial culture evolution. As well as, the research will produce objective measures of the deep learning network such as response time, and precision.

Self-evaluation: 100%

Research strategy

7/21. Methodological question: what is the qualitative, quantitative, or mixed-method of your research?

The research question is quantitative, we have prior data that we have to process, to develop a Deep learning algorithm, which will be evaluated with lab experiments and objective measures to characterize the system behavior. It is remarkable that qualitative methods are important since we are working with biological images, with high variability of shape, color, and mass but this analysis will be studied in a different paper.

Self-evaluation: 100%

Collection techniques

8/21. Methodological question: what are the data collection techniques of your research?

Our study will use TLM images from bacilli culture of mycobacteria tuberculosis. Every TLM set includes images from the same culture in 1-hour steps. The TML images include a channel of optical information in black and white. Additionally, The sequences

include another channel with the biomarker rRNA-GFP. The biomarker provides us the ability to differentiate bacilli metabolically active vs inactive bacterial [1].

[1] Manina, G., Dhar, N., & McKinney, J. D. (2015). Stress and host immunity amplify Mycobacterium tuberculosis phenotypic heterogeneity and induce nongrowing metabolically active forms. Cell host & microbe, 17(1), 32-46.

Self-evaluation: 100%

Analysis techniques

9/21. Methodological question: what are the data analysis techniques of your research?

The analysis of the data is quantitative. The data will be divided into a train, validation, and test set. Only the train set will be treated, the remaining sets will be salved for evaluation purposes. First, the train data will be analyzed in terms of the statistical importance and variance, this implies quantifying the number of bacteria per image and analyzing this distribution. Also, the number of unusual structures that appear in the images or the intensity variation will be evaluated. This leads us to understand the scope of the data. Then, The dataset will be cleaned, not adequate images will be deleted, like images not correctly taken, blurred, or without correct metadata reported. Then, the images will be used to develop de deep learning tool and the remaining validation and test sets will be used to characterize the tool behavior. All these procedures will be carried out in Python software.

Self-evaluation: 50%

Quality criteria

10/21. Methodological question: what are the tactics of your research to ensure scientific quality criteria?

To ensure the scientific quality criteria we use different types of validation. External validity, based on statistical generalization. Internal validity, the deep learning tool will be used and evaluated by microbiologist researchers in their diary work, and statistical measures of validity and reliability, doing several tests and evaluating parameters like accuracy, precision, and recall with random sampling carried out with Python software.

Self-evaluation: 50%

Unit of analysis

11/21. Empirical question: what is the unit of analysis of your research?

The unit of analysis is the interaction of the TLM images with the deep learning architecture. The result produced by the deep learning architecture will be compared with reality (microbiologist point of view), also the resources required to achieve the result will be analyzed.

Self-evaluation: 100%

Level of analysis

12/21. Empirical question: what is the level of analysis of your research?

The level of analysis is microscopic, as the TLM images, but the research has also tool-level analysis, the deep learning tool has to be analyzed in time and resources effectiveness, as well as, probabilistic measures like accuracy, recall, and precision.

Self-evaluation: 100%

Nature of data

13/21. Empirical question: what is the nature of the data of your research?

The nature of data is both qualitative and quantitative. Images are qualitative data, which implies a qualitative analysis in the preprocessing step, like discarding blurred images, foreign objects in the microscopic images... Also, the research has a lot of quantitative data like time of acquisition, information on drug testing, or growth conditions.

Self-evaluation: 100%

Origin of data

14/21. Empirical question: what is the origin of the data of your research?

The data is secondary data. We use real TLM images previously used by microbiologists for developing biomarkers and quantifying the number of living bacteria. This data will be now used for the first time to automatically detect and quantify the live bacteria using

deep learning. We expect new entrance of data, but always will be secondary, microbiologists take TML images to measure drug effectiveness, we expect to accelerate this measurement thanks to the research.

Self-evaluation: 100%

Sample

15/21. Empirical question: what is the sample of your research?

The current sample is a database of real TLM images composed of 52 sequences, that will be divided between train, validation, and test set. They were provided by microbiologists from The Institut Pasteur in Paris, with information about culture and image acquisition techniques. The TML includes mycobacterium tuberculosis treated with Isoniazid and pictures taken every hour to follow the behavior of the bacillus. The images are from 2013, but this database is expected to be increasing when the research starts to have positive results.

Self-evaluation: 100%

Pathos

16/21. Rhetorical question: what are the positive and negative emotions of your research?

The positive emotions of my research, are related to the public interest, we are working for accelerating drug development, to find new and more effective treatments, thanks to the automation of manual work. Furthermore, the automatization of manual work implies a lowering cost for pharmaceutical enterprises leading to commercial benefits.
This technology is faster and more accurate than manual work, but, we find negative emotions in the current workers doing this manual work, who feel the system is a threat, implying an ethical conflict between machine vs person.

Self-evaluation: 100%

Logos

17/21. Rhetorical question: what is the scientific logic of your research?

The logic of my work is abductive in applied research, we are researching a recent topic like deep learning, to discover and create a new technology to achieve the automatic recount and evaluation of microscopic bacilli.

Self-evaluation: 100%

Ethos

18/21. Rhetorical question: what are the limitations of your research?

During the research, the previous solutions are studied with a bibliographic review of the topic. Google Scholar shows about 16,900 scientific publications in 2022, which includes the main terms "deep learning" and "Microscopy". This big amount of publications leads to the main theoretical limitation of the research, being unable to cover all the theories and solutions presented in the literature.
The are also some empirical limitations, the training of the deep learning algorithm needs a great amount of data. This data has to be manually processed, and the amount of time required for this task limits the amount of information used during the study. Also, the study is limited computationally, the resources to compute the algorithm are finite, we want an algorithm efficient in time, GPU, and CPU.

Self-evaluation: 50%

Wisdom

19/21. Authorial question: what is your education and experience related to your research?

I am a telecommunication engineer, with a 4-years degree and a 2-years master, currently in the second year of my Ph.D. I have worked in the private sector, but my interest in the discovery of new solutions and techniques to help in the progress of medicine, lead me to specialize in artificial intelligence and start a research career.
In my final master's project, I worked in collaboration with the company eB2, a spin-off of Carlos III University, which designs and develops an eHealth web service for managing health information related to mental health and emotional well-being. Furthermore, to fully support my project, I have academic research experience. I have previously participated in the development of deep-learning algorithms for the detection of lung diseases in radiological images [1,2].

[1] Yang, D., Martinez, C., Visuña, L., Khandhar, H., Bhatt, C., & Carretero, J. (2021). Detection and analysis of COVID-19 in medical images using deep learning techniques. Scientific Reports, 11(1), 19638.

[2] Visuña, L., Yang, D., Garcia-Blas, J., & Carretero, J. (2022). Computer-aided diagnostic for classifying chest X-ray images using deep ensemble learning. BMC Medical Imaging, 22(1), 178.

Self-evaluation: 100%

Trust

20/21. Authorial question: who are the partners of your research?

This research is part of the European project ERA4tb (European Regimen Accelerator for Tuberculosis). The project is a public-private initiative devoted to accelerating the development of new treatment regimens for tuberculosis [1].
The ERA4TB Consortium brings together a multi-disciplinary team with proven expertise and capabilities in TB drug development to profile and progress anti-TB compounds.
The project consortium integrates 31 organizations, namely seven prestigious academic institutions (UC3M, UNIZAR, UU, EPFL, UHC, UNIPD, UPV), four non-profit organizations (IPP, IPL, iM4TB, BAR), nine public research organizations (FZB, CNR, CEA, SERMAS, PHE, NICE, SCI, IOS, CIM-Sant Pau) and five highly skilled small-medium enterprises (SYNAPSE, C-Path, IBT, QPS, GRIT), together with three EFPIA members (GSK, EVT, JANSSEN), and three IMI2 Associated Partners (BMGF, TBA, UNIVDUN).
The project ERA4tb allows the support of all the partners, this research maintains close contact with UC3M informatics researchers of the ARCOS group, which includes my Ph.D. supervisors and myself, and microbiologists researchers from UNIZAR.

[1] ERA4tb (European Regimen Accelerator for Tuberculosis). https://era4tb.org/

Self-evaluation: 100%

Time

21/21. Authorial question: what is your availability of time and resources for your research?

I´m working as a full-time researcher in the UC3M, with a predoctoral research scholarship, as part of the ERA4tb, a project with a big budget of 207MEuros. The project supports conference attendance and dissemination activities. This research is a part of my doctoral thesis, their development is the final part of my thesis and it's scheduled to develop in 6 months. The developed tool will be available for the whole project partners. We expect the continuity of the life of the research with updates and extensions. Furthermore, the tool can support the sustainability of the project as an exploitable result.

Self-evaluation: 50%