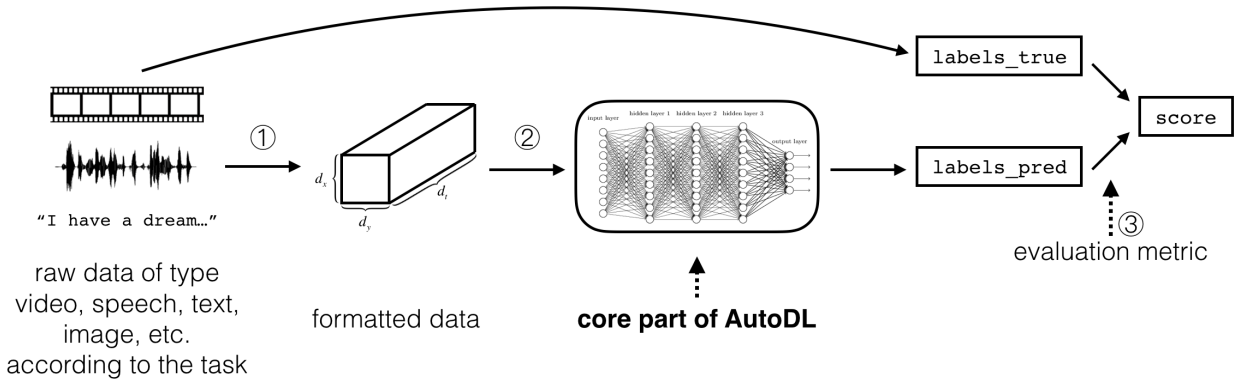
Title:

Automated Deep Learning Self-Service Against COVID-19

Topic: Machine Learning, Deep Learning
City and country: U. Paris-Saclay and INRIA (Gif-sur-Yvette) and Google (Zurich).
Team or project in the lab: Isabelle Guyon (U. Paris-Saclay and INRIA), Zhengying Liu (PhD. Student), Adrien Pavao (PhD. Student).
International collaboration with André Elisseeff (Google Zurich), Sergio Escalera (Univ. Barcelona), and Wei-Wei Tu (4Paradigm, China).
Name and mail of the principal advisors: Isabelle Guyon (iguyon@lri.fr), Zhengying Liu (zhengying.liu@lri.fr).
Department head: Joanne Cohen (jcohen@lri.fr) Université Paris-Saclay.

General presentation of the topic:



AI and Machine Learning are key technology that are currently being deployed against COVID-19 [1]. Of these technologies, "Deep Learning" is both one of the most effective and the hardest to harness. Deep Learning methods have known in the recent year tremendous success in applications, but also generate a lot of frustration from practitioners in need to quickly deploy them, particularly in a time when there is a shortage of experts. To address the latter problem, we have engaged in a series of AutoDL (Automated Deep-Learning) challenges with Google Zurich and 4Paradigm to foster research in the area of self-tuning Deep Learning learning algorithm (http://autodl.chalearn.org), which are part of the official selection of the prestigious NeurIPS conference. For these challenges, we have formatted nearly 100 datasets coming from a variety of domains and including speech, images, videos, text, and tabular data, all formatted in a unified way, but NOT preprocessed into fixed-length feature vectors. The first AutoDL challenge just ended and the winners have open-sourced their code [7-9]. Hence there is a great opportunity to make this code available to others and, in particular to lab that must quickly analyze COVID-19 data and may have a shortage of skilled engineers.

Objective of the internship:
The aim of this internship will be to help the challenge organizers in harvesting the results of the AutoDL challenge and specifically:
1) **Example of COVID data analysis**: Format one or several COVID datasets with https://github.com/zhengying-liu/autodl-contrib in the common tensor format used in the challenge, to serve as illustrative examples. Run them on the winners' solution and assess the results.
2) **AutoDL self-service:** Build a competition on the Codalab platform which, instead of having datasets/tasks waiting to be solved by algorithms will have winning algorithms (e.g. DeepWisdom, DeepBlueAI, etc.) waiting for users to upload their datasets. The algorithms will therefore be executed on the users' dataset(s) and provide without human intervention whatsoever machine learning model, ready-to-use. Elaborate on the feed-back provided by the code of the winners and the learning curve computed automatically by Codalab by designing an automatically generated human-legible report, suitable for a non-machine learning expert.
3) **AutoDL workflow:** Study the papers describing the winning solutions (currently in preparation). Extract a generic methodology and define a common principled workflow. This is the more academic part of the internship.

The intern will join a dynamic and motivated team of challenge organizers and may contribute to a publication on the applications of AutoDL to COVID-19.

Bibliographic references:
[1] I Mapping the landscape of Artificial Intelligence applications against COVID-19. Joseph Bullock, et la. 25 March, 2020. https://arxiv.org/pdf/2003.11336.pdf.

[2] Liu Z, Bousquet O, Elisseeff A, Escalera S, Guyon I, Jacques J, Pavao A, Silver D, Sun-Hosoya L, Treguer S, Tu WW. AutoDL Challenge Design and Beta Tests-Towards automatic deep learning. InCiML workshop@ NIPS2018 2018 Dec. https://hal.archives-ouvertes.fr/hal-01906226. Accessed October 2, 2019.

[3] B. Zoph, and L. Quoc. "Neural architecture search with reinforcement learning." arXiv preprint arXiv:1611.01578(2016).

[4] Elsken, Thomas, Jan Hendrik Metzen, and Frank Hutter. "Neural architecture search: A survey." arXiv preprint arXiv:1808.05377 (2018).

[5] Zhengying Liu, Isabelle Guyon, Julio Jacques,Jr, Meysam Madadi, Sergio Escalera, Adrien Pavao, Hugo Jair Escalante, Wei-Wei Tu, Zhen Xu, Sebastien Treguer, AutoCV Challenge Design and Baseline Results. https://hal.archives-ouvertes.fr/hal-02265053

[6] Madrid, J.G., Escalante, H.J., Morales, E.F., Tu, W.W., Yu, Y., Sun-Hosoya, L., Guyon, I. and Sebag, M., 2019. Towards AutoML in the presence of Drift: first results. arXiv preprint arXiv:1907.10772.

[7] Liu Z, Xu Z, Escalera S, et al. Towards Automated Computer Vision: Analysis of the AutoCV Challenges 2019. To appear in *Pattern Recognition Letters* of Elsevier. 2020. https://hal.archives-ouvertes.fr/hal-02386805. Accessed December 6, 2019.

[8] Liu Z, Xu Z, Rajaa S, Madadi M. Towards Automated Deep Learning: Analysis of the AutoDL challenge series 2019. To appear in *NeurIPSCD2019* in Proceedings of Machine Learning Research (PMLR) 2019:10.

[9] AutoDL challenge design and results (NAS workshop @ ICLR 2020) [SLIDES] (Talk in Google drive good resolution OR [Youtube part I] [Youtube part II]) [Website]

Expected ability of the candidate:
We are accepting candidates with background in machine learning, statistics, scientific modeling, signal processing and control (or at least a subset if those). The candidate should have the ability of working on cross-disciplinary problems, have a strong math background, and the experience or strong desire to work on practical problems. Good programming skills are also required. Experience with TensorFlow and/or PyTorch is preferred. Experience in GPU/TPU computing is a plus.