Dealing with Scarce Labelled Data: Semi-supervised Deep Learning with Mix Match for Covid-19 Detection Using Chest X-ray Images

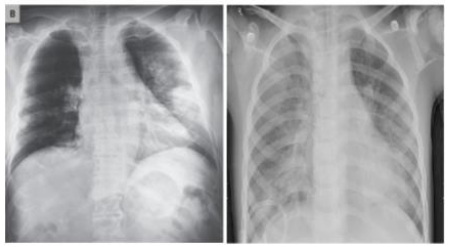
UNIVERSIDAD DE MÁLAGA

Saúl Calderón Ramírez, PhD student

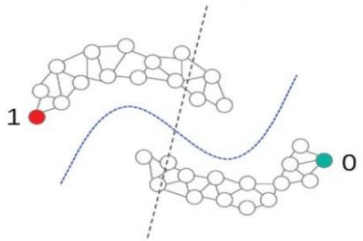
Introduction

Coronavirus is now a global pandemic, thus the detection of infected subjects in an early, quick and cheap manner is urgent

Deep learning for **Covid-19 detection using X-ray images** is an attractive approach, however it requires large labelled datasets



To overcome this challenge, a **semi-supervised deep learning** model using both labelled and unlabelled data is proposed, with the MixMatch architecture to classify chest X-rays into Covid-19, pneumonia and healthy cases



We also introduce a **semi-supervised deep learning boost coefficient** aimed to ease performance comparison of semi-supervised architectures

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Proposed method

- Previous work on COVID-19 detection using deep learning need large datasets and use data augmentation/ transfer learning to deal with scarce labelled datasets, with no literature on the usage of semi-supervised learning

- We evaluate the usage of MixMatch for COVID-19 detection, which proposes to minimize the following loss function in any CNN architecture:

$$\mathcal{L}\left(S, \boldsymbol{w}\right) = \sum_{\left(\boldsymbol{x}_i, \boldsymbol{y}_i\right) \in S'_l} \mathcal{L}_l\left(\boldsymbol{w}, \boldsymbol{x}_i, \boldsymbol{y}_i\right) + \gamma \sum_{\left(\boldsymbol{x}_j, \widetilde{\boldsymbol{y}}_j\right) \in \widetilde{S}'_u} \mathcal{L}_u\left(\boldsymbol{w}, \boldsymbol{x}_j, \widetilde{\boldsymbol{y}}_j\right)$$

- Where the first labelled term uses the augmented labelled dataset and the second one uses the unlabelled augmented dataset, which is obtained using the MixUp augmentation procedure, which creates linear combinations of the data:

$$\left(S'_l, \widetilde{S}'_u\right) = \Psi_{\mathrm{MixUp}}\left(S_l, \widehat{S}_u, \alpha\right)$$

Proposed method

- For the unlabelled term, MixMatch measures the consistency of the model output and the pseudo-labels using unlabelled augmented data:

$$\mathcal{L}_u\left(\boldsymbol{w}, \boldsymbol{x}_j, \widetilde{\boldsymbol{y}}_j\right) = \left\|\widetilde{\boldsymbol{y}}_j - f_{\boldsymbol{w}}\left(\boldsymbol{x}_j\right)\right\|$$

- The pseudo-label is calculated as the average label of the model when transforming the input:

$$\widehat{\boldsymbol{y}}_j = \frac{1}{K}\sum_{\eta=1}^{K} f_{\overrightarrow{w}}\left(\Psi^{\eta}\left(\boldsymbol{x}_j\right)\right)$$

- The sharpened label has proven to improve the pseudo-label:

$$\widetilde{\boldsymbol{y}} = \frac{\widehat{y}_i^{1/T}}{\sum_j \widehat{y}_j^{1/T}}$$

Proposed method: SSDL performance measurement

To compare SSDL methods, a measurement including different aspects is important, taking into account the proportion of unlabeled and evaluation data, through the usage of the proposed

$$\rho_{\mathrm{lu}} = \frac{n_l}{n_u + n_l}$$

Ratio of labelled and unlabelled data

$$\rho_{\mathrm{le}} = \frac{n_v}{n_v + n_l}$$

Ratio of validation and labelled data

$$\Delta_\rho = \frac{\overline{a}_{\text{semi-supervised}} - \overline{a}_{\text{supervised}}}{(\rho_{\mathrm{le}} + \rho_{\mathrm{lu}})\, s_{\text{semi-supervised}}}$$

SSDL boost coefficient

Results

- The data used is openly available from Dr. Cohen's repository for COVID-19+ cases, and a dataset of Chinese patients for COVID-19- cases
- We tested different unlabelled term weights, with positive gains when few labels are available of up to 15%, using WideResnet model with data augmentation and transfer learning

TABLE I

SEMI-SUPERVISED LEARNING ACCURACY (MEAN AND STD.) USING MIX MATCH (MM) FOR DIFFERENT UNSUPERVISED COEFFICIENTS VS. A FULLY SUPERVISED MODEL (F.S). ALWAYS $\rho_{LU} = 1$ FOR THE FULLY SUPERVISED MODEL. THE SIXTH COLUMN DENOTES THE CONFIDENCE P-VALUE OF THE ACCURACY DIFFERENCE BETWEEN MIX MATCH AND THE SUPERVISED MODEL.

Number of labels/coefficients	Fully supervised	$\gamma = 1$	$\gamma = 100$	$\gamma = 200$	F.S vs. MM ($\gamma = 200$)	Δ_ρ ($\gamma = 200$)
25 ($\rho_{\text{le}} = 0.24, \rho_{\text{lu}} = 0.11$)	0.683 ± 0.056	0.808 ± 0.053	0.816 ± 0.051	0.829 ± 0.057	$p = 2.36e-04$	7.318
40 ($\rho_{\text{le}} = 0.33, \rho_{\text{lu}=0.17}$)	0.729 ± 0.048	0.828 ± 0.04	0.848 ± 0.048	0.846 ± 0.048	$p = 0.0016$	4.875
50 ($\rho_{\text{le}} = 0.39, \rho_{\text{lu}=0.21}$)	0.785 ± 0.046	0.834 ± 0.038	0.843 ± 0.047	0.843 ± 0.049	$p = 0.0163$	1.972
70($\rho_{\text{le}} = 0.47, \rho_{\text{lu}=0.3}$)	0.808 ± 0.046	0.848 ± 0.053	0.864 ± 0.039	0.858 ± 0.041	$p = 0.1155$	1.5838
100 ($\rho_{\text{le}} = 0.56, \rho_{\text{lu}=0.43}$)	0.851 ± 0.049	0.853 ± 0.033	0.856 ± 0.051	0.854 ± 0.047	$p = 0.5194$	0.0648
All-undersampled (229)	0.896 ± 0.035					
All-imbalanced (4468)	0.966 ± 0.003					

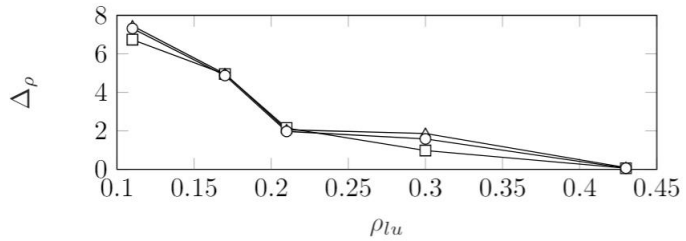


Fig. 4. Scalability curves using Δ_ρ against the ρ_{lu}. $\gamma = 200$ (triangle), $\gamma = 100$ (circle) and $\gamma = 1$ (square).

Hyper-parameter	Value
Image size	100×100 pixels
Weight decay	0.0001
Learning rate	0.00001
Batch size	12
Loss function	Cross entropy
Optimizer	Adam with 1-cycle policy [32]

Hyper-parameter	Description	Value
K	Number of augments	2
T	Sharpening temperature	0.5
α	Parameter for the dist. parameter	0.75

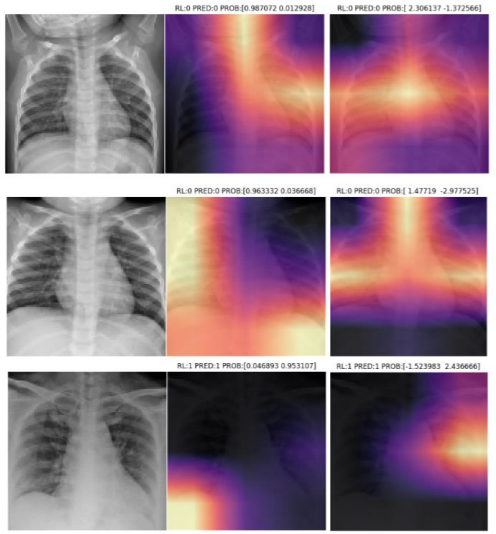


Fig. 3. From top to bottom: A three sample of the class activation maps for the tested dataset. From left to right: the original image, the heatmap of the usual supervised model, and the heatmap for the semi-supervised model. The legend RL corresponds to the real label, PRED to the model prediction and the array of two values is related to the output net values.

Conclusions and future work

- Unlabeled data using MixMatch can bring up to 15% of accuracy boost with very few labels, which can be useful for AI systems at the beginning of an outbreak
- Heatmaps in general showed to be more semantically relevant for the semi-supervised model
- The proposed semi-supervised accuracy boost coefficient enables the comparison of SSDL methods in practical applications, along a scalability analysis with different labelled/unlabelled data sample sizes
- Scalability problems for the tested MixMatch approach were revealed, with less gain using more labels
- Future work: analyzing the impact of SSDL in model robustness and uncertainty is important for medical applications, with additional specific transformations for radiological images

Thank you very much
Questions?

Email: sacalderon@itcr.ac.cr