An Attention Mechanism using Multiple Knowledge Sources for COVID-19 Detection from CT Images

Duy M. H. Nguyen [1,5], Duy M. Nguyen [2], Huong Vu [3], Binh T. Nguyen [4], Fabrizio Nunnari [1], Daniel Sonntag [1,6]

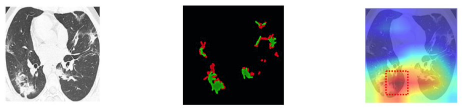
[1]German Research Center for Artificial Intelligence [2]School of Computing, Dublin City University, Ireland [3]University of California, Berkeley
[4]VNUHCM-University of Science, Vietnam [5]Max Planck Institute for Informatics, Germany, [6]Oldenburg University, Germany.

35-th AAAI Conference on Artificial Intelligence

Motivation

We propose a novel strategy to improve several performance baselines by leveraging multiple useful information sources relevant to doctors' judgments.

Infected regions and heat-map features extracted from learned networks are integrated with the global image via an attention mechanism during the learning process



(a) (b) (c)

Fig 1: (a) Picture of COVID-19 cases. (b) Red and green labels indicating the Ground-Glass Opacity and Pulmonary Consolidation regions. (c) The heat-mapregion extracted from the trained network

Method

Fusion with Multiple Knowledge

Infected branch

We extend the Semi Infected Net method from [1] to localize lung areas suffered by Ground-Glass Opacity and Pulmonary Consolidation on our CT images

Algorithm 1: Training Semi-supervised Infected Net

Input: D_{train} = D1 with segmentation masks and D_{test} = D2 ∪ D3 without masks.
Output: Trained Infected Net model, M

```
Set D_train = D1; D_test = D2 ∪ D3; D_subtest = NULL
while len(D_test) > 0 do
    Train M
    if len(D_test > 100) then
        D_subtest = random ( D_test \ D_subtest, k = 100)
        D_train = D_train ∪ M(D_subtest)
        D_test = D_test \ D_subtest
    else
        D_subtest = D_test
        M(D_subtest)
        D_test = D_test \ D_subtest
```

Heatmap branch

The heatmap H was extracted from the last convolution layer's output before computing the global pooling layer of the backbone.

This output was normalized across k channels as described in Eq. 1, where

- f_k is the activation unit
- $k = 1644$ (DenseNet169)
- $k = 2048$ (ResNet50)

The suspected regions B is the binarized mask of H with τ is the tuning parameter

$$H(x,y) = \frac{\sum_k f_k(x,y) - \min(\sum_k f_k)}{\max(\sum_k f_k)} \quad (1)$$

$$B = \begin{cases} 1, & \text{if } H(x,y) > \tau \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Multi-stream Network

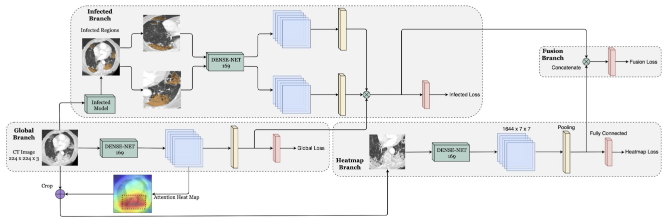


Fig 2: Our proposed attention mechanism with given a specific backbone network: infected regions (top branch), global image (middle branch), and learned heat-maps (bottom branch)

Input images at the ***Global Branch*** are fed into Infected-Model which is derived after completing the training procedure in Algorithm 1, to produce infected regions.

Heatmap regions from the global image extracted using Equation (1) and (2) are utilized as an input to train on **Heatmap Branch**.

Fusion Branch can be learned by merging all pooling layers from all branches.

Training Strategy

Due to limited amount of COVID-19 CT scans, it is not suitable to train entire branches simultaneously ⇒ Better to train each branch sequentially.

Sequential training method (Algorithm 2), where:

- W : trainable parameters
- I : input image of each branch
- ***g, h, in*** : global, heatmap, infected branch
- $Pool_f$: Pooling vector in **Fusion branch** by concatenating pooling vector of other branches

Algorithm 2: Training our proposed system

Input: Input image I_g, Label vector L, Threshold τ
Output: Probability score $p_f(c|I_g, I_h, I_{in})$

Learning W_g with I_g (**Stage I**);
Finding attention heat-map and its mapped image I_h of I_g by Eq. 2 and Eq. 1.
Learning W_h with I_h (**Stage II**);
Finding infected images I_{in} of I_g by using infected model M;
Learning W_{in} with I_{in} (**Stage II**);
Computing $Pool_f$, learning W_f, computing $p_f(c|I_g, I_h, I_{in})$ (**Stage III**).

Performance of training methods

Train the models with different training methods and different combination

- **GHIF** : train all branches together
- **GHI-F** : train the Fusion branch after training others branches at the same time
- **G-H-I-F** : train each branch as a sequence

Table 1: Accuracy of different training methods

Training	Global-Infected	Global-Heatmap	Fusion
GHIF	0.822	0.813	0.844
GHI-F	0.834	0.841	0.869
G-H-I-F	0.847	0.875	0.871

Experiments and Results

Dataset:

- **D1** [2]: *COVID-19 CT segmentation dataset*
  - 100 images from more than 40 COVID patients with lung segmentation and labeled infected area
- **D2** [1]: *COVID-19 CT Collection*
  - 1600 positive COVID images
- **D3** [3]: *Sample-Efficient COVID-19 CT Scans*
  - 349 positive and 397 negative COVID CT

Table 3: Performance on D3 with pretrained networks using self transfer techniques [3] combined with our pipeline

Method	Accuracy	F_1	AUC
ResNet50 *(Self-trans , Global)*	0.841	0.834	0.911
DenseNet169 *(Self-trans , Global)*	0.863	0.852	0.949
ResNet50 + *Our Infected*	0.842	0.833	0.918
ResNet50 + *Our heat-map*	0.879	0.848	0.924
ResNet50 + *Our Fusion*	0.861	0.870	0.927
DenseNet169 + *Our Infected*	0.853	0.849	0.948
DenseNet169 + *Our heat-map*	0.870	0.837	0.954
DenseNet169 + *Our Fusion*	0.882	0.853	0.964

Table 2: Performance on D3 with pretrained networks on ImageNet combined with our pipeline

Method	Accuracy	F_1	AUC
ResNet50 *(ImgNet, Global)*	0.803	0.807	0.884
DenseNet169 *(ImgNet, Global)*	0.832	0.809	0.868
ResNet50 + *Our Infected*	0.831	0.815	0.897
ResNet50 + *Our heat-map*	0.824	0.832	0.884
ResNet50 + *Our Fusion*	0.843	0.822	0.919
DenseNet169 + *Our Infected*	0.861	0.834	0.911
DenseNet169 + *Our heat-map*	0.855	0.825	0.892
DenseNet169 + *Our Fusion*	0.875	0.845	0.927

Table 4: Performance on D3 with SOTA network (Saeedi-Net[4] and Decaps[5]) combined with our pipeline

Method	Accuracy	F_1	AUC
Saeedi-Net	0.906 (±0.05)	0.901 (±0.05)	0.951 (±0.03)
Saeedi-Net + *Our w/o Semi*	0.913 (±0.03)	0.926 (±0.03)	0.960 (±0.03)
Saeedi-Net + *Our*	0.925 (±0.03)	0.924 (±0.03)	0.967 (±0.03)
Decaps $^{(1)}$	0.832 (±0.03)	0.837 (±0.03)	0.927 (±0.02)
Decaps $^{(1)}$+ *Our w/o Semi*	0.856 (±0.03)	0.864 (±0.03)	0.950 (±0.02)
Decaps $^{(1)}$+ *Our*	0.868 (±0.03)	0.872 (±0.03)	0.947 (±0.02)
Decaps $^{(2)}$	0.876 (±0.01)	0.871 (±0.02)	0.961 (±0.01)
Decaps $^{(2)}$+ *Our w/o Semi*	0.885 (±0.01)	0.884 (±0.02)	0.983 (±0.01)
Decaps $^{(2)}$+ *Our*	**0.896** (±0.01)	**0.889** (±0.01)	**0.986** (±0.01)

Interpretable Learned Features

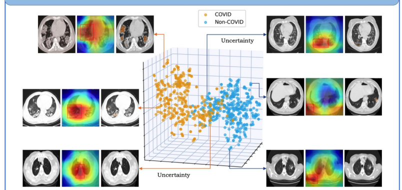


Fig 3: Interpreting learned features with the final layers of the fusion branch. Each point is presented together with its original scan, class activation map representation, and infected regions (left to right order)

Conclusion

Experiments showed that leveraging all visual cues yields improved performances of SOTA methods.

Our approach **provides more transparency** of the decision process to end-users by visualizing positions of attention maps, thereby increasing the model's interpretability in real-world applications.

References

[1] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji,Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and LingShao. Inf-net: Automatic covid-19 lung infection segmentation from ct images.IEEE Transactions on Medical Imaging, 2020

[2] https://medicalsegmentation.com/covid19/

[3] Xuehai He, Xingyi Yang, Shanghang Zhang,Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie.Sample-efficient deep learning for covid-19 diagnosis based on ct scans.medRxiv, 2020

[4] Abdolkarim Saeedi, Saeedi Maryam,and Arash Maghsoudi. A novel and reliable deep learning web-based tool to detect covid-19 infection from chest ct-scan.arXiv preprint arXiv:2006.14419, 2020.

[5] Aryan Mobiny, Pietro Antonio Cicalese, Samira Zare, Pengyu Yuan, Mohammadsajad Abavisani, Carol C Wu, Jitesh Ahuja, Patricia M de Groot, and Hien Van Nguyen. Radiologist-level covid-19 detection using ct scans with detail-oriented capsule networks.arXivpreprint arXiv:2004.07407, 2020.