CSE 6250 Student Project Proposal

Hua Zhao, M.S.
Georgia Tech

Motivation

Why is this problem important? Why do we care this problem? X-ray ¿ CT: CT should not beused to screen for or as a first-line test to diagnose COVID-19, CT should beused sparingly and reserved for hospitalized, symptomatic patients with specificclinical indications for CT https://arxiv.org/abs/2006.01174

Literature Survey

Conduct literature search to understand the state of arts and the gap for solving the problem. Formulate the data science problem in details (e.g., classification vs. predictive modeling vs.clustering problem).

- From https://github.com/ieee8023/covid-chestxray-dataset
  - Evaluation Issues 1. medical-irrelavent features extracted from images https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/
  - Disease Risks: https://www.youtube.com/watch?v=ineWmqfelEQ
    - Aatelectasis
    - Consolidation
    - Infiltration
    - ...
  - data software library: orchxrayvision
  - https://www.youtube.com/watch?v=nHQDDAAzIsI
    - dataset from kaggle for testing
    - first step: feature extraction: desease risks above? unknown? sliding window parsing 2d image? (convolutional + pooling) python: keras sequential model, fixed sliding window to extract feature?
    - second step: classify image based on extracted fine-grained images (denseNet) (CNN) can also try Transfer Learning model
    - segmentation techniques?
    - image classification models?
    - reasons for utilizing ML model
    - dataset from kaggle for testing
    - visualize extracted features (regions) on image: Grad-CAM, Saliency Maps, GSInquire
  - ROI (regions of interest):
    - Ground Glass Opacity (GGO)(belong to infiltrate)
    - Pulmonary consolidation.

Data

Describe the dataset you use, and elaborate onhow you would play with the data in your project.Preliminary results are encouraged but not required.It is recommended to try to cover as many aspectsas described in project initiation to give you a betternavigation in later period of project phase. This is acrucial step, please do it on the first day and neverstops until the project.

- data augmentation. (use a good PROTOCAL to merge different source datasets of images to avoid model learns meaningless patterns like two car images facing opposite directions, or in this case, for example, images uploaded by one doctor or from one hospital or from one X-ray machine has different CONTRAST than another set of images, or diff between CR and DX (computed/digital radiography). be careful not to increase unlikely data exisiting in real world)
  - flipm
  - rotation
  - scale
  - crop
  - translation (move objects in the image background by x,y axis)
  - gaussian noise - to distort high-frequency useless extracted features while also distort useful ones, may not be useful because model needs high-resolution image in some regions?
  - advanced: conditional GANs transfer overall 'conditions' of the image, like seasonality for a landscape imgae. computationally expensive.
  - advanced: neural style transfer (like Deep Photo Style Transfer) same goal as conditional GANs by mixing 'style' of one image using another, cheaper computation. cons: output is more artistic than realistic
  - interpolation (translate image beyond original boundary, likely not used) 1. constant 2. edge 3. reflect (useful for continuous or natural backgrounds like forests, grassland) 4. symmetric (similar to reflect) 5. wrap
  - contrast (NEED TO DEAL WITH THIS FOR CHEST XRAY IMAGE)
  - histtogram equalization =¿ segmentation

Approach

Identify the high-level technical approachesfor the project (e.g., what algorithms to use or pipelinesto use). Identify clearly the success metric that youwould like to use (e.g., AUC, accuracy, recall, speedupin running time). (Metric: AUC, accuracy, specificity, sensitivity, BA, DOR)

- CNN
- transfer learning: VGG19 network
- transfer learning: nanonet (imbed augumentation, not open source)
- COVID-net?

Experimental Setup

Setup the analytic infrastructurefor your project (including both hardware and softwareenvironment, e.g., AWS or local clusters with Spark,python and all necessary packages).

Timeline

prepare a timeline and milestones of deliv-erables with reasonably proposed task distributions forthe entire project.

References

1. Pryor TA, Gardner RM, Clayton RD, Warner HR. The HELP system. J Med Sys. 1983;7:87-101.
2. Gardner RM, Golubjatnikov OK, Laub RM, Jacobson JT, Evans RS. Computer-critiqued blood ordering using the HELP system. Comput Biomed Res 1990;23:514-28.