AI and medical imaging

Camille Ruppli, PhD - Data Scientist
Overview

- Introduction and context
- Addressing the lack of annotations: self-supervised learning approaches
- Application to prostate cancer detection
Introduction and context

- Medical examinations: images we are now quite familiar with
- First MRI in France: 1984
- Since then: drastic increase in images amount

Source: https://data.oecd.org/fr/healthcare/
Introduction and context

- Data increase along with development of deep learning methods
- Natural images: ImageNet, CoCo, Places

Gap of images and easy annotation amount

**Introduction and context**

**Classification**
- Healthy / Pathological
- Chest pathology
- < 2 minutes

**Localization**
- Box around region of interest
- ~ 5/10 minutes

**Segmentation**
- Pixel-wise contour of pathology or organ
- Up to 30 minutes
Addressing the lack of annotations: supervised transfer learning

- **IMAGENET**
- **model $f_\theta$**
  - Trained from scratch
  - Fine-tuned weights with $\theta_0 = \theta$
- **model $f_\theta^*$**
  - 70% Dog
  - 10% Cat
  - ... 0.1% Flower
  - 90% Alzheimer
  - 10% Healthy
Introduction and context

Domain gap

- Visually dissimilar
- Medical images: smaller differences between classes
- Medical images can be 3D

Introduction and context

Large unlabeled / healthy database

Small annotated cohort

Pretext task

Transfer weights

90% Alzheimer
10% Healthy

Trained from scratch

Fine-tuned weights with $\theta_0 = \theta$
Self-supervised learning approaches

Generation based self-supervised methods

- Transformation types
- Link with target task

<table>
<thead>
<tr>
<th>encoder $f_\theta$</th>
<th>decoder $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer weights</td>
<td></td>
</tr>
</tbody>
</table>

model $f_\theta$:
- 90% Pathological
- 10% Healthy

Trained from scratch

Fine-tuned weights with $\theta_0 = \theta$
Self-supervised learning approaches

Contrastive learning methods

Similarity loss
Most frequent cancer and cause of death from cancer in men

Estimated number of incident cases and deaths Europe, males, all ages

- Prostate
- Lung
- Colorectum
- Bladder
- Kidney
- Stomach
- Melanoma of skin
- Pancreas
- Non-Hodgkin lymphoma
- Liver

Data source: GLOBOCAN 2020
Graph production: Global Cancer Observatory (http://gco.iarc.fr/)
© International Agency for Research on Cancer 2023
Application to prostate cancer detection

Morphology: T2w

Diffusion

Diffusion coefficient

Prostate

Prostate lesion
Application to prostate cancer detection

PI-RADS score
- lesion malignancy level, from 1 to 5
- High annotator variability

Figure from Sonn et al, European Urology Focus, 2019
Application to prostate cancer detection

- Access to multi-modality MRI
- Prostate segmentation
  - Easier task, less annotator variability [1]
- Manual lesion segmentation
  - More costly to obtain
  - More variability

Application to prostate cancer detection

- Patient metadata also available
  - Different types
  - Not available for every patient
  - Subject to annotator variability: PI-RADS scores

→ Contrastive learning approaches to take advantage of unannotated data and available metadata

- Lesion PIRADS score
- Patient’s age
- Radiological report
- ISUP / Gleason scores

- Patient age
- PSA level
- Global PIRADS score from radiological report

- Patient’s age
- Radiological report

Patient’s metadata
Application to prostate cancer detection

- Assumption: two patients with similar metadata $p$ (e.g. PI-RADS score) should be close in the representation space
- Our contribution: including PI-RADS variability and confidence

Application to prostate cancer detection

- Reference segmentation
- Predicted lesion
Conclusion

- Advent of neural networks on natural images with increased data availability
- Annotations much more complex to obtain in medical domain
- Self-supervised approaches: addressing the lack of annotations
- Promising results on prostate cancer detection taking variable metadata into account