Deep Learning for Life Science

Hossein Azizpour
Assistant Professor in Machine Learning
Robotics, Perception, and Learning (RPL)
KTH
The time is right!
necessary technological developments come together

- CMOS
- Protein & chemical labels
- miniaturization & high throughput
- tissue slide scanners
- automation
- SSDs, network storage
- Cloud computing
- Faster CPUs
- GPUs
- deep learning

- More sensitive labels and sensors
- Higher throughput data creation
- Increased storage capacity
- Faster computing
- Better algorithms

- high throughput confocal
- cryo-EM
- super resolution
- light sheet

- More advanced microscopy techniques
- Protein & chemical labels
- miniaturization & high throughput
- tissue slide scanners
- automation
- SSDs, network storage
- Cloud computing
- Faster CPUs
- GPUs
- deep learning

- The time is right!
necessary technological developments come together

- Bioimage Informatics
A New Era
of large publicly available data and benchmarks

<table>
<thead>
<tr>
<th>Year(s)</th>
<th>Challenge</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-2010</td>
<td>DIADEM</td>
<td>Neuron morphology</td>
</tr>
<tr>
<td>2012</td>
<td>PTC</td>
<td>Particle detection in time-lapse microscopy</td>
</tr>
<tr>
<td>2012</td>
<td>SNEMI23</td>
<td>Neurite segmentation in 2D</td>
</tr>
<tr>
<td>2012,2014</td>
<td>MITOS</td>
<td>Cell mitosis detection in histopathology</td>
</tr>
<tr>
<td>2012,2014</td>
<td>DM3D</td>
<td>3D deconvolution microscopy</td>
</tr>
<tr>
<td>2013-2015</td>
<td>CTC</td>
<td>Cell segmentation and tracking</td>
</tr>
<tr>
<td>2013</td>
<td>SMLM</td>
<td>Single-molecule localization</td>
</tr>
<tr>
<td>2013</td>
<td>SNEMI3D</td>
<td>Neurite segmentation in 3D</td>
</tr>
<tr>
<td>2013</td>
<td>AMIDA</td>
<td>Cell mitosis detection in histopathology</td>
</tr>
<tr>
<td>2014,2015</td>
<td>OCCIS</td>
<td>Overlapping cell segmentation in cancer</td>
</tr>
<tr>
<td>2014</td>
<td>MITOS-ATYPIA</td>
<td>Mitosis detection and cancer scoring</td>
</tr>
<tr>
<td>2015</td>
<td>GLAS</td>
<td>Gland segmentation in histopathology</td>
</tr>
<tr>
<td>2015</td>
<td>NCC</td>
<td>Nucleus counting in multichannel micro.</td>
</tr>
<tr>
<td>2015-2016</td>
<td>BigNeuron</td>
<td>Large-scale 3D neuron reconstruction</td>
</tr>
<tr>
<td>2016-2017</td>
<td>CAMELYON</td>
<td>Cancer metastasis detection</td>
</tr>
<tr>
<td>2017</td>
<td>DREAM</td>
<td>Breast cancer detection</td>
</tr>
<tr>
<td>2017</td>
<td>CYTO</td>
<td>Cell atlas protein localization</td>
</tr>
</tbody>
</table>
A New Era

Excitement surrounding bioimage informatics
A New Era of AI-assisted medical advances
A New Era of AI-assisted medical advances

Accuracy

Lack of specialists
A New Era of AI-assisted medical advances

21 Board Certified Stanford Dermatologists
129,450 images of 2,032 diseases
1.41 million AI training images

Accuracy

Lack of specialists

Efficiency
A New Era of AI-assisted medical advances

- **Accuracy**

- **Efficiency**

- **Lack of specialists**

- **Cheaper**
A New Era
General Life Science

SciLifeLab

THE HUMAN PROTEIN ATLAS

A Tissue-Based Map of the Human Proteome

A large number of tools can be accessed with direct links to gene-specific images of the corresponding proteins in the different tissues and organs.

verily

AstraZeneca

H₂N

SO₂

N

Ag⁺
Contents

• Problem Definition

• Opportunities and challenges

• An example pipeline

• Domain Adaptation

• Uncertainty Estimation

• Future Directions
Contents

• Problem Definition

• Opportunities and challenges

• An example pipeline

• Domain Adaptation

• Uncertainty Estimation

• Future Directions
Deep Learning for Medical Science
Deep Learning for Medical Science

Supervised Learning
\[ D \sim P(X,Y) \] \& \[ P(X) \] \quad \[ P(Y|X) \]
- classification
- regression

Under-supervised Learning
\[ D \sim P(X,Y) \] \& \[ P(X) \] \quad \[ P(Y|X) \] or \[ P(Z|X) \]
- weakly supervised learning
- semi-supervised learning

Un/self Supervised Learning
\[ D \sim P(X) \]
- clustering
- dimensionality reduction

Discriminative Models
Generative Models
Deep Learning for Medical Science

- Prognosis
- Risk Estimation
- Diagnosis
- Novel Biomarker
- Patient Care
- New Treatments
- Personalized Medicine
Contents

• Problem Definition

• Opportunities and challenges

• An example pipeline

• Domain Adaptation

• Uncertainty Estimation

• Future Directions
Opportunities

• Large amount of data is available (stored somewhere, hopefully digitized)
  • High-content screening
  • Digital pathology
  • Radiology
  • Drug discovery
  • Genomics
  • Clinical data
  • ...

• Repetitive routine tasks
• Shortage of specialists, outsourcing diagnosis/prognosis to different countries
• Training procedure for new specialists
• Improved and/or consistent accuracy
• Efficient diagnosis process
• Interesting scientific challenges
• Private and public Investments
• Last but not least, direct impact on human lives
Challenges

- Reliability
  - Interpretability
  - Uncertainty
- Recall is extremely important, precision also very important
- Public scrutiny for automated systems and AI in general
- Data standardization
- Different imaging equipment with different internal parameters and operator settings
- Multi-modal data
- Non-Euclidean data
- Bias (regional, temporal, etc.). Fair machine learning.
- Annotation cost (specialists’ time are precious/expensive)
- Ambiguous definitions → Disagreement among specialists
- Privacy Issues with data collection
- Reverse engineering networks
- Inherently noisy labels
- Deep learning is data-hungry
Challenges

- **Reliability**
  - Interpretability
  - Uncertainty
- Recall is extremely important, precision also very important
- Public scrutiny for automated systems and AI in general
- Data standardization
- **Different imaging equipment with different internal parameters and operator settings**
  - Multi-modal data
  - Non-euclidean data
  - Bias (regional, temporal, etc.). Fair machine learning.
  - Annotation cost (specialists’ time are precious/expensive)
  - Ambiguous definitions → Disagreement among specialists
- Privacy Issues with data collection
- Reverse engineering networks
- Inherently noisy labels
- Deep learning is data-hungry
Contents

• Problem Definition
• Opportunities and challenges
• An example pipeline
• Domain Adaptation
• Uncertainty Estimation
• Future Directions
Medical Projects

- Histopathology
- Radiology
- Standard camera
Medical Projects

- Histopathology
- **Radiology** ✓
- Standard camera
Digital mammographic screening for breast cancer

Current screening in Sweden

Women invited for screening*

Digital mammograms acquired from multiple angles

Two radiologists review case

Discussion and decision

Healthy

Recall

1. Additional imaging
2. Physical exam
3. Tissue sample

Cancer diagnosis

It is estimated that mammography-based screening reduces breast cancer mortality by around 30%.

*All women in Sweden between 40 and 74 are invited for screening every 18 to 24 months.
Critical shortage of radiologists

"Akut personalbrist hotar mammografin"

Landstingen klarar inte själva att lösa dilemmat med att landets mammografiinruter dröneras på personal, samtidigt som belastningen ökar. Nu krävs konkret handling för att trygga mammografiutredningen, skriver Ulrika Årehed Kåström och Jan Zedenius, Cancerfonden.
Limitations in Computer aided detection (CAD)

Can Computer-aided Detection Be Detrimental to Mammographic Interpretation?

Mammographic interpretation is one of the most difficult tasks in the field of radiology. Reading mammograms can be more of an art than a science. Breast parenchymal patterns are not stable between patients, between left and right breasts, and even within the same breast from year to year in the same patient. Possessing, particularly in the mediolateral-oblique projection, and the amount of compression applied are variable from examiner to examiner. Breast cancer has a varied appearance on mammograms. From dense to amorphous, to very subtle anomalies noted on only one view. In fact, the conditions were real with full digital resolution or a magnifying glass. The large volume of cases requiring interpretation in many practices is also daunting, given the number of women in the population for whom yearly screening mammography is recommended. Task repetition and fatigue contribute to making the subtle signs of breast cancer a very real possibility, but not that much more common.

Given these difficulties, it is not surprising that approximately 20% of cancers are never to be missed at mammography (1-7). Some of this is because of the inherent sensitivity of mammography in the detection of lesions in dense breast tissues, but even if a lesion is visible on the images, the resolution of the mammographic view is limited by the inherent nature of mammographic absorption, as well as the intervening radiographic thickness due to both detecting and deciding to act on it. recall) such lesions, affects the reading. Accuracy in mammographic interpretation depends on many factors, of which experience and scheme of analysis are never distinct that this difficult task could likely be made less error prone with the help of computer programs. However, how such is art and how much is vs. even a computer is expected to operate in a very important manner (a variable mammography and differing breast cancer pattern) at least this would be clearly relevant (is not making or many cancer without causing more problems or, many false positive results)? It is a lot to ask, even of a computer.

CAD Studies and How They Fail

The literature about CAD is concerned with various factors influencing its validity and validity of results. There are some important differences in the CAD studies that have been published. While this does not negate the results, it does throw into question the validity of some studies. Results of some of the early studies, which were performed by using retrospective analyses or computer modeling, suggested that CAD can achieve the same or better results for this. However, in the cancer detection rate of radiologists with the use of CAD, there was actually a great deal of improvement over the concept of CAD and the results of these early studies.

Assessment of the early studies was based on retrospective analyses. That is, CAD was applied to mammograms with known cancers, and the physicians reviewed to determine if CAD identified the cancers. Such studies were performed in an artificial environment. Due to these, the detection of the CAD system was not expected to exceed the radiologists' performance, that is, to be more accurate in identifying cancers. However, in a recent study, the results were quite different. The CAD system identified more cancers than the radiologists, as shown in the figure below.

Clinical success depends on CAD having a high sensitivity, a reasonable specificity, and the reader taking appropriate action when interpreting the CAD prompts. All of these features work in conjunction, and all need to be optimized for the system to be valuable. Because mammography is an imperfect system, particularly in the detection of many of the more subtle cancers in denser breast tissues, to have reasonable sensitivity, the false-positive rate of CAD prompts must be high; thus, the specificity is low. Thus, the balance is not easy to achieve.

Difficult cases result in unacceptably high FP rate
Optimized Workflow

Women invited for screening

Detectability estimate
Or
Cancer Risk Estimate

60% → Radiologist or CAD

30% → Two radiologists review

10% → One radiologist reviews

10% and/or Cancer Risk Estimate.

Discussion & decision

MRI and/or

One specialist reviews

Decision

*or cancer risk estimate.
Massive dataset:

Screening: 241,149 exams
Clinical: 57,752 exams
Screening register:
~400,000 women (entire SLL)
Cancer register:
~9,000 women (entire SLL)

Total: ~2,400,000 images
Cancer: ~20,000 images

INbreast
Total: 410 images
Cancer: 360 images

CBIS-DDSM
Total: 4,067 images
Cancer: 855 images
Deep Learning for Mammography

Tumor Detection Network (CAD) and Risk Estimation
Tumor detection

Based on state-of-the-art Inception-ResNet\(^1\) architecture

Training the network

Modified the pretrained network (ImageNet) so it is **fully convolutional**

First, trained the network to localize tumors using existing annotated datasets (semi-supervised learning)

![Database icons](image1.png)

- **Dream**
  - Total: 500 images
  - Cancer: 32 images

- **CBIS-DDSM**
  - Total: 4,067 images
  - Cancer: 855 images

- **INbreast**
  - Total: 410 images
  - Cancer: 360 images

**Data augmentation – generated millions of training examples from a few thousand**

![Data augmentation methods](image2.png)

- random crops
- aspect ratio
- rotation
- contrast
- brightness
Risk Estimation

- Risk: risk of a woman developing cancer at some point in future
- Positive set: prior and contra-lateral mammograms

<table>
<thead>
<tr>
<th>Method</th>
<th>AUC</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Learning Risk</td>
<td>64%</td>
<td>5.32 (2.39-9.69)</td>
</tr>
<tr>
<td>Mammographic Density Risk</td>
<td>57%</td>
<td>1.96 (1.23-3.11)</td>
</tr>
</tbody>
</table>

Figure 1: Representations of the 4 Breast Imaging Reporting and Data System (BI-RADS) breast density qualitative and quantitative assessments. A) BI-RADS 1: almost entirely fat; B) BI-RADS 2: scattered fibroglandular densities; C) BI-RADS 3: heterogeneously dense; and D) BI-RADS 4: extremely dense.

[Dembrower, Azizpour, Smith, Konuk, Strand “TRAINING A DEEP LEARNING NETWORK TO ASSESS BREAST CANCER RISK”. , CARS 2018]
Contents

• Problem Definition

• Opportunities and challenges

• An example pipeline

• Domain Adaptation

• Uncertainty Estimation

• Future Directions
Knowledge Transfer

$$D_s \sim P_s(X, Y) \quad M_s \equiv P_s(Y_s|X) \text{ or } P(X,Y_s)$$

**SOURCE**

Data → Model

Rich/large

$$D_t \sim P_t(X, Y) \quad M_t \equiv P_t(Y_t|X) \text{ or } P(X,Y_t)$$

**TARGET**

Data → Model

small/noisy
Transfer Learning

• when $Y_s \neq Y_t$ or $P_s(Y_s|X)$
Deep Transfer Learning

- Fine tuning is the most common way
Deep Transfer Learning

[Azizpour et al. Factors of Transferability for a Generic ConvNet Representation, PAMI 2016]
Deep Transfer Learning

Width vs Depth

[Azizpour et al. Factors of Transferability for a Generic ConvNet Representation, PAMI 2016]
Deep Transfer Learning

Which Layer?

[Azizpour et al. Factors of Transferability for a Generic ConvNet Representation, PAMI 2016]
# Deep Transfer Learning

## Multiple Representations

<table>
<thead>
<tr>
<th>Source task</th>
<th>Image Classification</th>
<th>Attribute Detection</th>
<th>Fine-grained Recognition</th>
<th>Compositional</th>
<th>Instance Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VOC07</td>
<td>MIT</td>
<td>SUN</td>
<td>H3D</td>
<td>UIUC</td>
</tr>
<tr>
<td>ImageNet</td>
<td>71.6</td>
<td>64.9</td>
<td>49.6</td>
<td>73.8</td>
<td>90.4</td>
</tr>
<tr>
<td>Places</td>
<td>68.5</td>
<td>69.3</td>
<td>55.7</td>
<td>68.0</td>
<td>88.8</td>
</tr>
<tr>
<td>Hybrid</td>
<td>72.7</td>
<td>69.6</td>
<td>56.0</td>
<td>72.6</td>
<td>90.2</td>
</tr>
<tr>
<td>Concat</td>
<td>73.8</td>
<td>70.8</td>
<td>56.2</td>
<td>74.2</td>
<td>90.4</td>
</tr>
</tbody>
</table>

[Azizpour et al. Factors of Transferability for a Generic ConvNet Representation, PAMI 2016]
Deep Transfer Learning

Diversity vs Density

[Azizpour et al. Factors of Transferability for a Generic ConvNet Representation, PAMI 2016]
Deep Transfer Learning

Optimizing Transferability Factors are Important!

<table>
<thead>
<tr>
<th>Image Classification</th>
<th>Attribute Detection</th>
<th>Fine-grained Recognition</th>
<th>Compositional</th>
<th>Instance Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-ConvNet</td>
<td>[38]</td>
<td>[25]</td>
<td>[44]</td>
<td>[30]</td>
</tr>
<tr>
<td>Deep Standard</td>
<td>71.1</td>
<td>68.5</td>
<td>37.5</td>
<td>87.5</td>
</tr>
<tr>
<td>Deep Optimized$^4$</td>
<td>80.7</td>
<td>71.3</td>
<td>56.0</td>
<td>92.5</td>
</tr>
<tr>
<td>Err. Reduction</td>
<td>32%</td>
<td>18%</td>
<td>13%</td>
<td>13%</td>
</tr>
</tbody>
</table>

| Source Task | ImgNet | Hybrid | Hybrid | ImgNet | ImgNet | ImgNet | ImgNet | ImgNet | Hybrid | ImgNet | ImgNet | ImgNet | Hybrid | ImgNet | ImgNet | ImgNet | ImgNet |
| Network Width | Medium | Medium | Medium | Medium | Large | Medium | Medium | Medium | Medium | Medium | Medium | Medium | Medium | Medium | Medium | Medium | Medium | Medium |
| Network Depth | 16 | 8 | 8 | 8 | 16 | 16 | 16 | 16 | 16 | 16 | 8 | 8 | 16 | 16 | 16 | 16 |
| Rep. Layer | last | last | last | 2nd last | 2nd last | 2nd last | 3rd last | 3rd last | 3rd last | 3rd last | 4th last | 4th last | 4th last | 4th last | 4th last | 4th last | 4th last |
| PCA | X | X | X | X | X | X | X | X | X | X | ✓ | ✓ | ✓ | ✓ | ✓ |
| Pooling | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | 1 × 1 | 1 × 1 | 2 × 2 | 2 × 2 | 3 × 3 |

[Azizpour et al. Factors of Transferability for a Generic ConvNet Representation, PAMI 2016]
Domain Adaptation

- when $P_S(X) \neq P_t(X)$
Domain Adaptation

DDSM

INBREAST

DREAM
Domain Adaptation

There is usually the assumption that samples drawn from $P_s(X)$ have a corresponding sample in $P_t(X)$.

- Supervised Domain Adaptation
- Unsupervised Domain Adaptation

- Directly adapting the model parameters [Rozantsev et al. 2017]
- Learning a common embedding space [Ganin et al 2015, 2016]
- Adapting in the input domain [Bousmalis et al. 2016 Liu et al 2017]
Domain Adaptation

Domain Adaptation

\[
\min_{E_1, E_2, G_1, G_2} \ L_{VAE_1}(E_1, G_1) + L_{CC_1}(E_1, G_1, E_2, G_2) + V_{LSGAN}(E_1, G_1) \\
L_{VAE_2}(E_2, G_2) + L_{CC_2}(E_2, G_2, E_1, G_1) + V_{LSGAN}(E_2, G_2)
\]

[Liu et al. Unsupervised Image-to-Image Translation Networks, NIPS 2017]
Domain Adaptation

[Tao Wang, Adapting multiple datasets for better mammography tumor detection, KTH 2018]
Domain Adaptation

\[
\begin{align*}
\min_{E_1, E_2, G_1, G_2} & \quad L_{VAE_1}(E_1, G_1) + L_{CC_1}(E_1, G_1, E_2, G_2) + V_{LSGAN}(E_1, G_1) \\
& \quad + L_{VAE_2}(E_2, G_2) + L_{CC_2}(E_2, G_2, E_1, G_1) + V_{LSGAN}(E_2, G_2) \\
& \quad + L_{seg}(E_2, G_1) + L_{seg}(E_1, G_2)
\end{align*}
\]

[Tao Wang, Adapting multiple datasets for better mammography tumor detection, KTH 2018]
Domain Adaptation

[Tao Wang, Adapting multiple datasets for better mammography tumor detection, KTH 2018]
[Ronneberger et al. “U-Net: Convolutional Networks for Biomedical Image Segmentation” MICCAI 2015]
Domain Adaptation

[Tao Wang, Adapting multiple datasets for better mammography tumor detection, KTH 2018]
# Domain Adaptation

<table>
<thead>
<tr>
<th>Training dataset</th>
<th>Testing dataset</th>
<th>Pixel Level Metrics</th>
<th>Instance Level Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dice</td>
<td>Precision</td>
</tr>
<tr>
<td>CBIS</td>
<td>INBreast</td>
<td>0.1170</td>
<td>0.1965</td>
</tr>
<tr>
<td>CBIS</td>
<td>Transferred INBreast</td>
<td>0.1412</td>
<td>0.1544</td>
</tr>
<tr>
<td>Transferred CBIS</td>
<td>INBreast</td>
<td>0.1934</td>
<td>0.2521</td>
</tr>
</tbody>
</table>

[Tao Wang, Adapting multiple datasets for better mammography tumor detection, KTH 2018]
Domain Adaptation

Synthetic-to-Real Adaptation

[Image of a diagram showing the UNIT model with nodes labeled $E_S$, $G_S$, $x_S$, $G_T$, $x_T$, $D_S$, $D_T$, $x_{S \rightarrow T}$, $x_{T \rightarrow S}$, $x_{T \rightarrow T}$, $x_{S \rightarrow S}$, and $z$.

[Equation for the UNIT model with notations for reconstruction loss, adversarial loss, and ground-truth preservation loss.

Domain Adaptation

Synthetic-to-Real Adaptation

[Source image table with examples from SYNTHIA-Seq and SYNTHIA-RAND-CVPR16]

Domain Adaptation

Synthetic-to-Real Adaptation

[Image of synthetic-to-real adaptation comparison]

Contents

• Problem Definition
• Opportunities and challenges
• An example pipeline
• Domain Adaptation
• **Uncertainty Estimation**
• Future Directions
Why do we care about uncertainty?

**Sources**
- Model Uncertainty
  - Small Dataset
  - Covariate shift
- Input Uncertainty
  - Noisy measurements
  - Incomplete input

**Applications**
- Active Learning
- Exploration/Exploitation
- Risk associated with certain decisions
- Model Pathology
- Forecasting
- Non-Stationary Environment

And many more…
What do we mean by modelling uncertainty?

A regression example

\[ f_\theta(x) \in \mathbb{R}^d \times \mathbb{R} \]

Multi-modal distributions!
Bayesian Modeling
for epistemic uncertainty

\[ P(y|x, D) = \int P(y|x, D, \theta) P(\theta|D) \, d\theta \]

\[ P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \]
Uncertainty with Deep Networks

- Place distribution over model parameters [Denker&LeCun 1991, Neal 1995]
- Variational Approximation [Hinton&Camp 1993, Graves 2011, Blundell et al. 2015]
- Probabilistic Backpropagation [Rezenede et al. 2014, Lobato&Adams 2015]
  - .....

These approaches require redesigning the architecture and training procedure and are usually computationally expensive to train.
Uncertainty with Deep Networks
No change in the training procedure

• Monte Carlo Dropout [Gal & Ghahramani 2016]

• Batch Normalization Dropout [Azizpour, Teye, Smith 2018]

• Frequentist Uncertainty [Lakshminarayanan 2017]

• Other approaches (e.g. entropy of the softmax distribution)
Stochastic Regularization as VI

Stochastic Regularization Techniques (SRT)

as approximate Bayesian inference

[Gal & Ghahramani, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, JMLR 2016]
DropOut

\[ \theta_i = \begin{cases} 
  w_i & \text{with prob } p \\
  0 & \text{with prob } 1 - p 
\end{cases} \]

[Gal&Ghahramani, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, JMLR 2016]
Batch Normalization

\[ \overline{w_i} = \frac{w_i - \mu_B}{\sigma_B} \]

Toy Dataset

Quantitative Results
Regression - Normalized Measure

\[ \sum_{i=1}^{N} \log P(\hat{y}_i = y_i) \]

Lower bound variance: \( \tau^{-1} \)
Upper bound variance: \( \sigma^* \)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>PLL MCBN</th>
<th>PLL MCDO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston Housing</td>
<td>10.49****</td>
<td>5.51****</td>
</tr>
<tr>
<td>Concrete</td>
<td>-36.36**</td>
<td>10.92****</td>
</tr>
<tr>
<td>Energy Efficiency</td>
<td>10.89****</td>
<td>-14.28*</td>
</tr>
<tr>
<td>Kinematics 8nm</td>
<td>1.68***</td>
<td>-0.26 ns</td>
</tr>
<tr>
<td>Power Plant</td>
<td>0.33**</td>
<td>3.52****</td>
</tr>
<tr>
<td>Protein Tertiary Structure</td>
<td>2.56****</td>
<td>6.23****</td>
</tr>
<tr>
<td>Wine Quality (Red)</td>
<td>0.19*</td>
<td>2.91****</td>
</tr>
<tr>
<td>Yacht Hydrodynamics</td>
<td>45.58****</td>
<td>-41.54 ns</td>
</tr>
</tbody>
</table>

Quantitative Results
Classification - CIFAR

\[ \sum_{i=1}^{N} \log P(\hat{y}_i = y_i) \]

Baseline (standard network with dataset-average BN): \(-0.32\)

<table>
<thead>
<tr>
<th>Number of stochastic forward passes</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLL</td>
<td>-0.36</td>
<td>-0.32</td>
<td>-0.30</td>
<td>-0.29</td>
<td>-0.29</td>
<td>-0.28</td>
<td>-0.28</td>
<td>-0.28</td>
</tr>
</tbody>
</table>

Qualitative Results

Error plot

Qualitative Results
Semantic Segmentation - CamVid

Qualitative Results
Semantic Segmentation – Pascal VOC

Qualitative Results
Semantic Segmentation – Pascal VOC

Simple Recipe

- Take any pretrained network with batch normalization and/or dropout layers
- Calculate constant observation noise on training/validation set, call it $\tau^{-1}$
- At test time, sample different batches and/or dropout masks, get the predictions set
- Calculate the mean and standard deviation of the predictions, $\mu, \sigma$
- The new point estimate of our prediction is: $\mu$
- The associated uncertainty to it is: $\tau^{-1} + \sigma^2$

Conclusion

• Positive points
  • Standard training
  • Simple algorithm
  • Vast applicability

• Negative points
  • Lots of assumptions
  • Under/over estimating the uncertainty
  • Considerable computation at test time

Contents

• Problem Definition
• Opportunities and challenges
• An example pipeline
• Domain Adaptation
• Uncertainty Estimation
• Future Directions
Future Directions

- Geometric Deep Learning
- Side information (clinical information) as well as privileged information
- In a general scale: Multi-modal learning (unformatted textual description, images, clinical information, sequencing data)
- Probabilistic Uncertainty
- Weakly supervised learning (finding patterns that specialists are not aware of) new biomarkers, would make big leaps in life science
- Causality
- With missing data
- Encrypted networks
- Learning with noisy labels
Questions