How Important Is Each Dermoscopy Image?

Catarina Barata and Carlos Santiago
Motivation

Dermoscopy Datasets
Motivation

Deep Networks Like Data
Motivation

Class Distribution

Dermoscopy Datasets

- MEL
- NV
Motivation

Class Distribution

Dermoscopy Datasets

- MEL
- NV
Motivation

Class Distribution

Dermoscopy Datasets

BKL
MEL
NV

PH2
ISIC'16
ISIC'17
Motivation

Class Distribution

Dermoscopy Datasets
Motivation

Class Distribution

Dermoscopy Datasets

- SCC
- VASC
- DF
- BCC
- AKIEC
- BKL
- MEL
- MEL
- NV
Motivation

Why is this a problem?

• Network bias

• Poor Generalization

<table>
<thead>
<tr>
<th>Class</th>
<th># Samples</th>
<th>Deep Net Recall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NV</td>
<td>6741</td>
<td>95</td>
</tr>
<tr>
<td>MEL</td>
<td>1119</td>
<td>66</td>
</tr>
<tr>
<td>BKL</td>
<td>1101</td>
<td>77</td>
</tr>
<tr>
<td>AKIEC</td>
<td>331</td>
<td>45</td>
</tr>
<tr>
<td>BCC</td>
<td>517</td>
<td>88</td>
</tr>
<tr>
<td>DF</td>
<td>116</td>
<td>43</td>
</tr>
<tr>
<td>VASC</td>
<td>143</td>
<td>68</td>
</tr>
</tbody>
</table>
Me Likes Balanced Data More...
Challenges

- Deal with class imbalance
- Not all classes are equally hard
- Are all samples equally important?

<table>
<thead>
<tr>
<th>Class</th>
<th># Samples</th>
<th>Deep Net Recall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NV</td>
<td>6741</td>
<td>95</td>
</tr>
<tr>
<td>MEL</td>
<td>1119</td>
<td>66</td>
</tr>
<tr>
<td>BKL</td>
<td>1101</td>
<td>77</td>
</tr>
<tr>
<td>AKIEC</td>
<td>331</td>
<td>45</td>
</tr>
<tr>
<td>BCC</td>
<td>517</td>
<td>88</td>
</tr>
<tr>
<td>DF</td>
<td>116</td>
<td>43</td>
</tr>
<tr>
<td>VASC</td>
<td>143</td>
<td>68</td>
</tr>
</tbody>
</table>
Goal

How to make the most of the available data?

• Data Augmentation
• Importance Sampling
• Sample Weighting
Goal

How to make the most of the available data?

• Data Augmentation

• Importance Sampling

• Sample Weighting
Sample Weighting

Batch Samples

DNN

Loss

Feedforward and Compute Sample Loss $\ell_j$

Backpropagate and Update Model Parameters

$\mathcal{L} = \frac{1}{M} \sum_{j}^{M} \ell_j$
Sample Weighting

Batch Samples → Feedforward and Compute Sample Loss $\ell_j$ → Loss

- Cross Entropy Loss (CEL)
- Focal Loss (CEL)

Loss:

$$\mathcal{L} = \frac{1}{M} \sum_{j} \ell_j$$

Backpropagate and Update Model Parameters
Sample Weighting

Batch Samples → DNN → Sample Weights → Loss

Feedforward and Compute Sample Loss $\ell_j$

$$\mathcal{L} = \frac{1}{M} \sum_{j} w_j \ell_j$$

Backpropagate and Update Model Parameters
Weighting Strategies

• Class-Balanced Losses
  – Class-balanced (CB)\(^1\):
    \[
    w_j = \frac{N}{N y_j}
    \]
  – Effective Number of Samples (ES)\(^2\):
    \[
    w_j = \frac{1 - \beta}{1 - \beta^{N y_j}}, \quad \beta = \frac{N - 1}{N}
    \]

\(^1\) Provost, *Machine Learning From Imbalanced Datasets* 101, AAAI 2000
\(^2\) Cui et al., *Class-balanced Loss Based on Effective Number of Samples*, CVPR 2019
Weighting Strategies

- **Class-Balanced Losses**
  - Class-balanced (CB):
    \[ w_j = \frac{N}{N_{y_j}} \]
  - Effective Number of Samples (ES):
    \[ w_j = \frac{1-\beta}{1-\beta N_{y_j}}, \quad \beta = \frac{N-1}{N} \]
Weighting Strategies

• Curriculum Learning

\[
\arg\min_w \frac{1}{M} \sum_{j} w_j \ell_j + G(w; \lambda)
\]
Weighting Strategies

• Curriculum Learning
  – Self-paced Learning (SPL)\textsuperscript{[1]}:
    \[ G(w; \lambda) = -\lambda \|w\|_1 \]
  – Online Hard Example Mining (OHEM)\textsuperscript{[2]}:
    \[ G(w; \lambda) = +\lambda \|w\|_1 \]

\textsuperscript{[1]} Kumar et al., Self-Paced Learning for Latent Variable Models, NeurIPS 2010
\textsuperscript{[2]} Shrivastava et al., Training Region-based Object Detectors with Online Hard Example Mining, CVPR 2016
Weighting Strategies

• Curriculum Learning
  – Self-paced Learning (SPL):
    \[ G(w; \lambda) = -\lambda \|w\|_1 \]
  – Online Hard Example Mining (OHEM):
    \[ G(w; \lambda) = +\lambda \|w\|_1 \]
Experimental Setup

- DNN Architectures
  - Flat Classifier (VGG-16)
  - Hierarchical Classifier\textsuperscript{[1]}
- Dataset
  - ISIC 2018
- Performance Metrics
  - Recall
  - Precision
  - F1-Score
  - Accuracy
  - Balanced Accuracy

\textsuperscript{[1]} Barata et al., Explainable Skin Lesion Diagnosis Using Taxonomies, Pattern Recognition 2020
\textsuperscript{[2]} Woo et al., CBAM: Convolutional Block Attention Module, ECCV 2018

+ CBAM\textsuperscript{[2]}
## Results

<table>
<thead>
<tr>
<th>Loss</th>
<th>Acc</th>
<th>BAcc</th>
<th>mPR</th>
<th>mF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEL</td>
<td>87.5</td>
<td>75.5</td>
<td>80.0</td>
<td>77.0</td>
</tr>
<tr>
<td>FL</td>
<td>87.0</td>
<td>74.5</td>
<td>79.0</td>
<td>76.0</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Loss</th>
<th>Acc</th>
<th>BAcc</th>
<th>mPR</th>
<th>mF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEL</td>
<td>87.5</td>
<td>75.5</td>
<td>80.0</td>
<td>77.0</td>
</tr>
<tr>
<td>CB</td>
<td>84.0</td>
<td>78.4</td>
<td>76.0</td>
<td>76.5</td>
</tr>
<tr>
<td>ES</td>
<td>84.5</td>
<td>76.7</td>
<td>77.0</td>
<td>76.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Loss</th>
<th>Acc</th>
<th>BAcc</th>
<th>mPR</th>
<th>mF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>87.0</td>
<td>74.5</td>
<td>79.0</td>
<td>76.0</td>
</tr>
<tr>
<td>CB</td>
<td>83.0</td>
<td>76.9</td>
<td>73.0</td>
<td>75.0</td>
</tr>
<tr>
<td>ES</td>
<td>83.5</td>
<td>78.0</td>
<td>74.0</td>
<td>75.5</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Loss</th>
<th>Acc</th>
<th>BAcc</th>
<th>mPR</th>
<th>mF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEL</td>
<td>-</td>
<td>87.5</td>
<td>75.5</td>
<td>80.0</td>
</tr>
<tr>
<td></td>
<td>CB</td>
<td>84.0</td>
<td>78.4</td>
<td>76.0</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>84.5</td>
<td>76.7</td>
<td>77.0</td>
</tr>
<tr>
<td></td>
<td>SPL</td>
<td>85.5</td>
<td>68.8</td>
<td>76.5</td>
</tr>
<tr>
<td></td>
<td>OHEM</td>
<td>87.0</td>
<td>76.4</td>
<td>79.0</td>
</tr>
<tr>
<td>FL</td>
<td>-</td>
<td>87.0</td>
<td>74.5</td>
<td>79.0</td>
</tr>
<tr>
<td></td>
<td>CB</td>
<td>83.0</td>
<td>76.9</td>
<td>73.0</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>83.5</td>
<td>78.0</td>
<td>74.0</td>
</tr>
<tr>
<td></td>
<td>SPL</td>
<td>84.5</td>
<td>65.3</td>
<td>71.5</td>
</tr>
<tr>
<td></td>
<td>OHEM</td>
<td>88.0</td>
<td>75.7</td>
<td>80.5</td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>CEL</th>
<th>FL</th>
<th>CEL-ES</th>
<th>CEL-SPL</th>
<th>CEL-OHEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>VASC</td>
<td>VASC</td>
<td>VASC</td>
<td>VASC</td>
<td>VASC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VASC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCC</td>
<td>BCC</td>
<td>BCC</td>
<td>DF</td>
<td>BCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>CEL</th>
<th>FL</th>
<th>CEL-ES</th>
<th>CEL-SPL</th>
<th>CEL-OHEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEL</td>
<td>MEL</td>
<td>MEL</td>
<td>NV</td>
<td>MEL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCC</td>
<td>AKIEC</td>
<td>AKIEC</td>
<td>AKIEC</td>
<td>AKIEC</td>
</tr>
</tbody>
</table>

MEL: Melanoma
AKIEC: Actinic Keratoses and Intradermal Carcinoma

Images show the results of different computer vision techniques applied to skin lesions.
Conclusions

• Weighting strategies significantly affect the performance of a DNN
• Some weighting schemes may induce bias
• Features learned by DNNs change according to the learning strategy
• OHEM achieves the best overall performance
Thank You!

ana.c.fidalgo.barata@tecnico.ulisboa.pt