Al-cyclopedia: A look into fascinating art and understanding why they are what they are

Shaoyang Luo
Roman Puttkammer
Kana Sethu

{sluo3, rputtkam, ksethu}@Stanford.edu
IMET2021FGVC8 Dataset

- Train: 145K images
- Test: 53K
- Unique Attributes: 3475
Histogram of attributes

- Most images have 6
- But some can have as much as 20!
- Many of which can be conflicting
Image Classification with DNN

Experiments
- 3 CNN Architectures
- Metrics: PR-AUC, F2
  - BCE LOSS

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Train</th>
<th>Validation Weighted PR-AUC</th>
<th>Test PR-AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ViT</td>
<td>.95</td>
<td>.59</td>
<td>.59</td>
</tr>
<tr>
<td>Simple CNN</td>
<td>.40</td>
<td>.40</td>
<td>.39</td>
</tr>
<tr>
<td>EfficientNET</td>
<td>.41</td>
<td>.37</td>
<td>.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>F2 Threshold=0.5</th>
<th>F2 Threshold=OPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ViT</td>
<td>.23</td>
<td>.28</td>
</tr>
<tr>
<td>Simple CNN</td>
<td>.14</td>
<td>.15</td>
</tr>
<tr>
<td>EfficientNET</td>
<td>.11</td>
<td>.16</td>
</tr>
</tbody>
</table>
(b) Smooth Grad

(d) Guided Backprop
Attribute: medium::glaze
Future Work

• Self-supervised pretraining of ViTMAE to improve latent representations.
• Alternate/improved data augmentation.
• ViT-specific methods for saliency.

(See project report for all references.)
NP-Hard

NP-Complete

NP

P

Complexity

P ≠ NP

P = NP

P = NP

P ≠ NP

P = NP
Thank You!