GLICO Training

- Let \( \{x_i\}_{i=1}^{n} \) denote a set of labeled images, choose \( n \) \( d \)-dimensional random learnable vectors on the unit sphere \( [x_i], x_i \in S^d \).
- Pair every image \( x_i \) with a random vector \( z_i \), to achieve the mapping \( (x_i, z_i) \). Learn the parameters \( \theta \) of the generator \( G \) and the optimal set \( \{z_i\} \) by minimizing the objective:
  \[
  \min_{\theta, \{z_i\}} \mathbb{E}_{x \sim \mathcal{D}} \mathbb{E}_{e \sim \mathcal{D}} \mathcal{L}_{\text{rec}}(G(x,e), z_i) + \mathcal{L}_{\text{CE}}
  \]

- Train a classifier \( F \) to classify the labeled data \( \{x_i, y_i\} \) when available:
  \[
  \min_{\theta} \mathbb{E}_{x \sim \mathcal{D}} \mathcal{L}_{\text{CE}}(F(G(x,e)), y_i)
  \]

- \( \mathcal{L}_{\text{CE}} \) is the cross-entropy loss.

Conceptual Illustration

- Illustration of the latent space \( Z \).
  - (a) Naïve Generative Latent Optimization (GLO): Vectors \( z_i \neq Z \) do not have semantic meaning in \( Z \) space.
  - (b) Our method: Vectors from the same class are grouped.
  - (c) Our method in transductive mode.

Notations: Filled colored circles represent different labeled datapoints, where color corresponds to class identity. Black circles with the symbol "?" represent unlabeled datapoints.

Experimental Results

- For each benchmark, we defined a small sample task by subsampling the original training set of the corresponding dataset.
- Classification NN backbone are WideResNet-28 for CIFAR-10 and CIFAR-100, and ResNet50 for CUB-200.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Samples</th>
<th>Class</th>
<th>Top-1 Acc.</th>
<th>Top-5 Acc.</th>
<th>Top-2 Acc.</th>
<th>Top-3 Acc.</th>
<th>Top-4 Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIFAR-100</td>
<td>10</td>
<td>22.69</td>
<td>28.55</td>
<td>40.42</td>
<td>32.26</td>
<td>40.21</td>
<td>23.01</td>
</tr>
<tr>
<td>25</td>
<td>28.39</td>
<td>39.60</td>
<td>52.05</td>
<td>43.31</td>
<td>51.62</td>
<td>44.31</td>
<td>30.01</td>
</tr>
<tr>
<td>50</td>
<td>47.82</td>
<td>59.56</td>
<td>67.41</td>
<td>57.11</td>
<td>64.01</td>
<td>54.11</td>
<td>44.11</td>
</tr>
<tr>
<td>100</td>
<td>61.73</td>
<td>64.27</td>
<td>74.09</td>
<td>67.31</td>
<td>71.64</td>
<td>64.01</td>
<td>55.09</td>
</tr>
<tr>
<td>CUB-200</td>
<td>5</td>
<td>50.79</td>
<td>51.22</td>
<td>69.18</td>
<td>59.63</td>
<td>69.18</td>
<td>57.60</td>
</tr>
<tr>
<td>10</td>
<td>64.11</td>
<td>64.11</td>
<td>80.08</td>
<td>72.11</td>
<td>82.12</td>
<td>70.08</td>
<td>62.12</td>
</tr>
<tr>
<td>20</td>
<td>69.11</td>
<td>69.11</td>
<td>80.08</td>
<td>72.11</td>
<td>82.12</td>
<td>70.08</td>
<td>62.12</td>
</tr>
<tr>
<td>30</td>
<td>73.15</td>
<td>73.15</td>
<td>80.08</td>
<td>72.11</td>
<td>82.12</td>
<td>70.08</td>
<td>62.12</td>
</tr>
</tbody>
</table>

Table: Top-1 accuracy (%) including STE with a different number of training samples per class (labeled data only).

References

[1] B. Barz and J. Szita, "Deep learning on small datasets without pre-training using clever initialization" 2020

Small Sample Learning ≠ Few Shot Learning

- The Small Sample settings are substantially different from the two related settings of Semi-Supervised Learning (SSL) and the Few-Shot (FS) learning.
- FS- The learner has access to many labeled examples from classes not participating in the current classification task. Thus, most FS algorithms rely on transfer learning from tens of thousands of labeled training examples.
- SSL- The learner typically has access to many unlabeled examples. Most SSL algorithms transfer knowledge from the distribution of the unlabeled data.

Image Classification Using GLICO

- Train the proposed model GLICO as described above by minimizing the sum of the reconstruction loss and the cross-entropy loss:
  \[
  \min_{\theta} \mathcal{L}_{\text{rec}}(G(x,e)) + \mathcal{L}_{\text{CE}}
  \]
- Train a classifier as follows:
  - Sample two pairs \( (x, z), (x, z) \in (X, X) \)
  - where \( z, x, z \in C \), \( C \subset C, C \) is the classes set
  - \( \mathcal{L}_{\text{inter}} = \mathcal{L}_{\text{ECE}}([z, z, x]), r \sim U[0, 0.4] \)
  - Alternate with probability \( p \) ≈ 0.5 training inputs for the classifier \( \hat{G}(\hat{\mathcal{D}}_{\text{inter}}(J)) \) and the image \( x \),

- **Spherical linear interpolation**