1. Motivation

- Data imbalance is common in visual classification

Wearing hat
Not wearing hat

Minority class
Majority class

2. Main Idea

- Deep embedding: Class-level → cluster- & class-level constraint
  - Triplet embedding
  - Quintuplet embedding

3. Large Margin Local Embedding (LMLE)

- Triple-header hinge loss

\[
\min \sum_i (\varepsilon_i + r_i + s_i) + \lambda |W|_2^2
\]

\[
\max (0, g_i + D(f(x_i), f(x_i^{*''})) - D(f(x_i), f(x_i^{*''''}))) \leq \varepsilon_i
\]

\[
\max (0, g_i - D(f(x_i), f(x_i^{*''})) - D(f(x_i), f(x_i^{*''''}))) \leq r_i
\]

\[
\max (0, g_i + D(f(x_i), f(x_i^{*''})) - D(f(x_i), f(x_i^{*''''}))) \leq s_i
\]

\[\forall i, \varepsilon_i \geq 0, r_i \geq 0, s_i \geq 0\]

- Network architecture

  - Equal class re-sampling & class costs assignment in batches

4. Cluster-wise kNN search

- Large margin cluster-wise kNN: fast & imbalance resistant

\[
y_i = \arg \max_{y \in \{1, \ldots, C\}}\min_{y' \neq y}\frac{D(f(q), f(m_{y'})) - \max_{y' \neq y}}{D(f(q), f(m_y))}
\]

5. Results

- Large-scale CelebA face attribute dataset
  - 200K celebrity images, each with 40 attributes
  - Highly imbalanced: average positive class rate 23%
  - Total accuracy = \frac{\text{tp}}{\text{tp} + \text{fn}} \text{ Balanced accuracy } = \frac{\text{tp}(p + in)}{(\text{tp} + \text{fn})}

<table>
<thead>
<tr>
<th></th>
<th>Total acc.</th>
<th>Balanced acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triplet-kNN</td>
<td>83</td>
<td>72</td>
</tr>
<tr>
<td>Aneal</td>
<td>87</td>
<td>80</td>
</tr>
<tr>
<td>LMLE-kNN</td>
<td>90</td>
<td>84</td>
</tr>
</tbody>
</table>

- Edge detection on BSDS500 dataset

  - Retrieve from 2M edge label patches with long-tail distribution

6. Conclusion

- Cluster- & class-level quintuplets preserve both locality across clusters and discrimination between classes, irrespective of class imbalance
- Large margin classification by fast cluster-wise kNN search