Learning Deep Representation for Imbalanced Classification

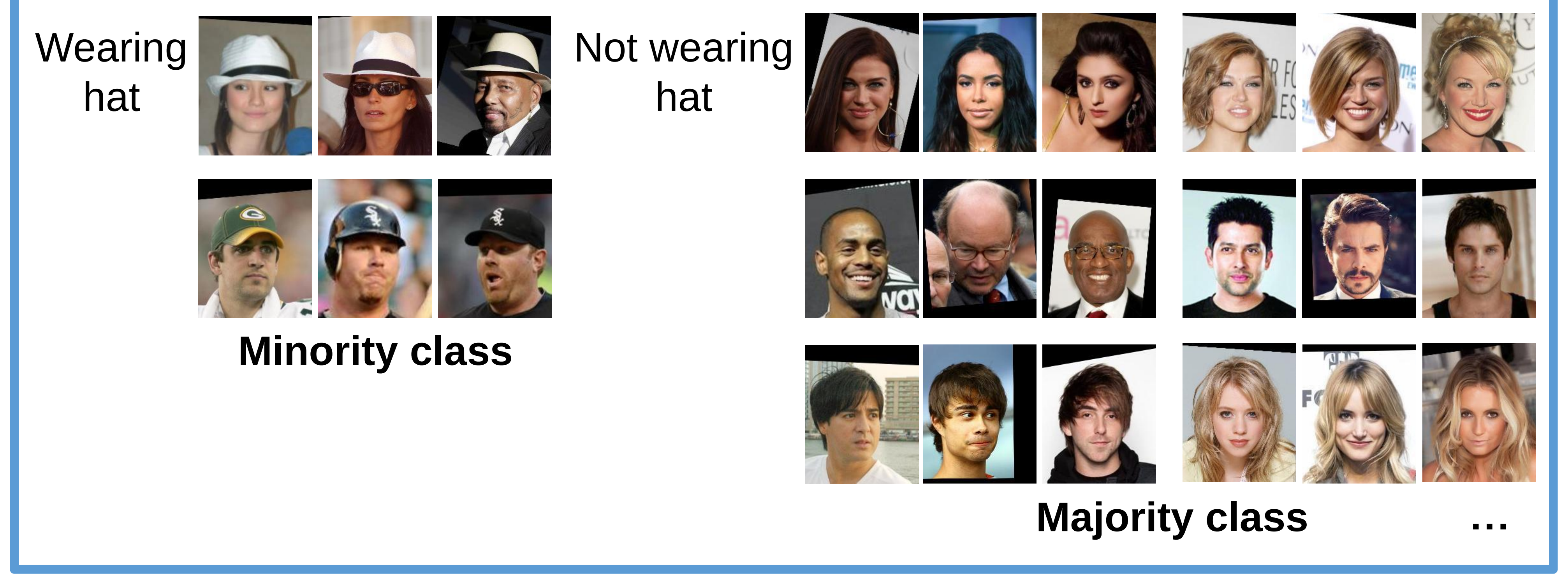
Chen Huang[1,2], Yining Li[1], Chen Change Loy[1], Xiaoou Tang[1]
[1]The Chinese University of Hong Kong [2]SenseTime Group Limited
{chuang, ly015, ccloy, xtang}@ie.cuhk.edu.hk

CVPR 2016

1. Motivation

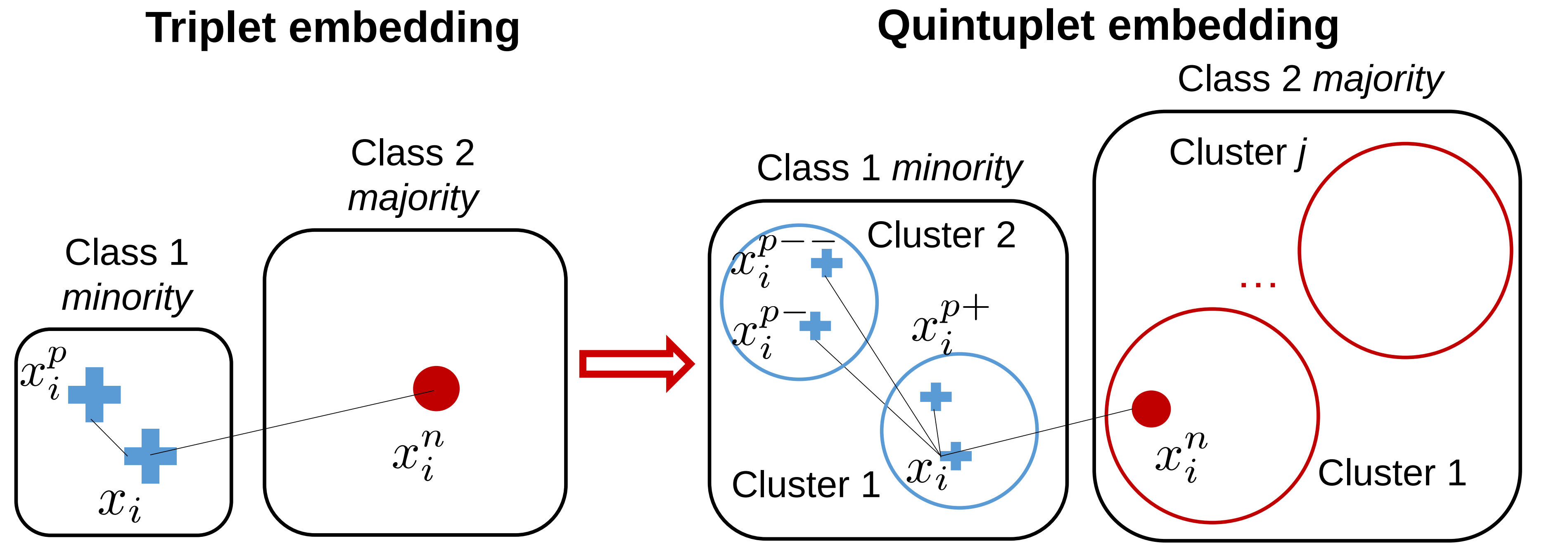
- Data imbalance is common in visual classification

Face attribute example



2. Main Idea

- Deep embedding: **Class-level** → **cluster- & class-level** constraint



$$D\left(f(x_i),f(x_i^{p+})\right) < D(f(x_i),f(x_i^{p-}))$$
$$< D(f(x_i),f(x_i^{p--})) < D(f(x_i),f(x_i^{n}))$$

x_i – an anchor
x_i^{p+} – the anchor's most distant within-cluster neighbor
x_i^{p-} – the nearest within-class neighbor of the anchor, but from a different cluster
x_i^{p--} – the most distant within-class neighbor of the anchor
x_i^{n} – the nearest between-class neighbor of the anchor

- Study traditional **re-sampling** and **cost-sensitive learning** scheme

3. Large Margin Local Embedding (LMLE)

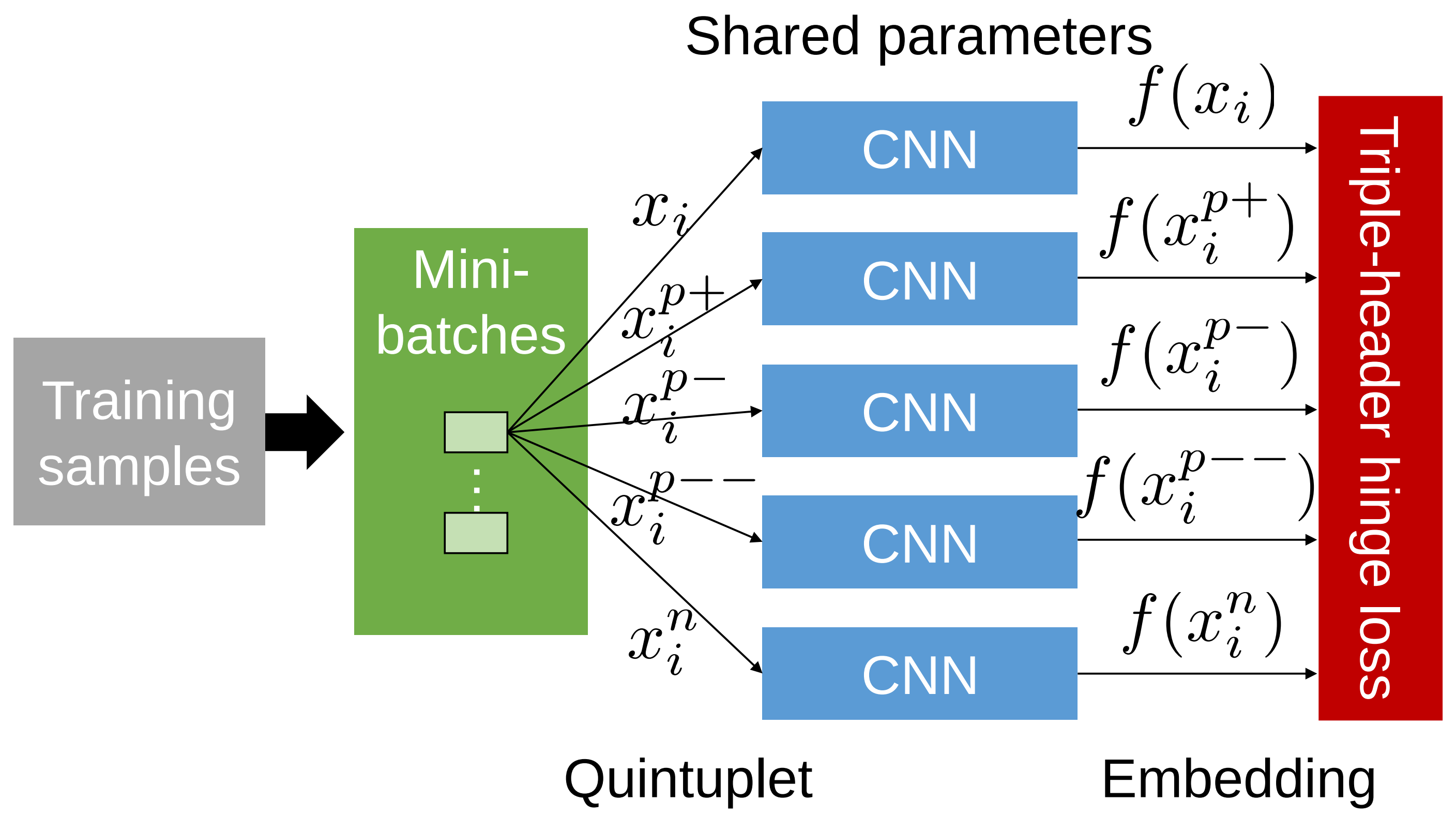
- Triple-header hinge loss

$$\min \sum_i (\varepsilon_i+\tau_i+\sigma_i)+\lambda\|\mathbf{W}\|_2^2$$

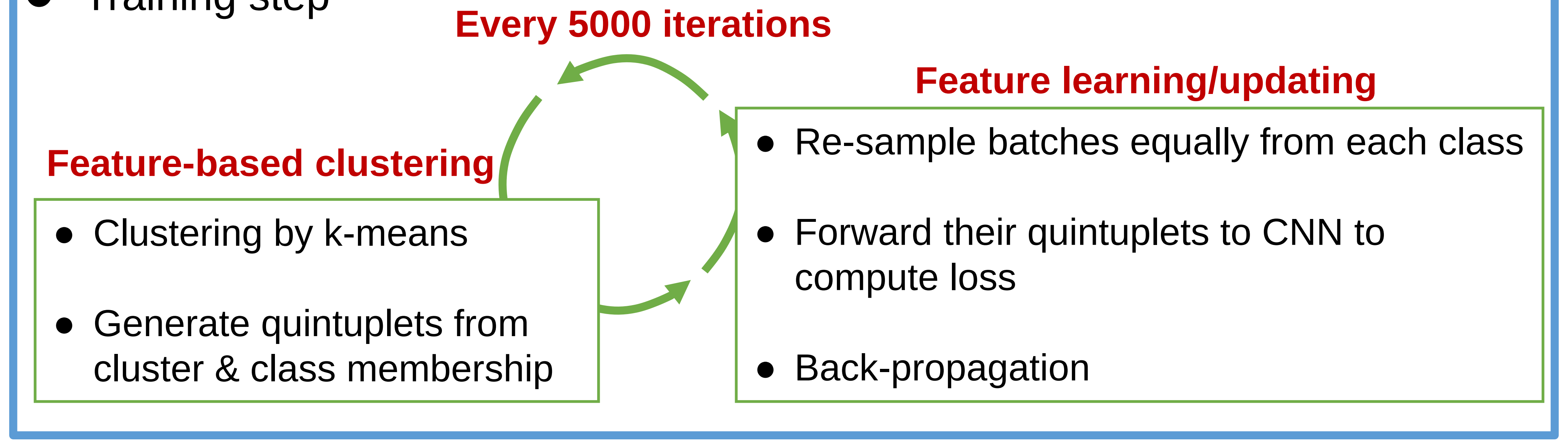
s.t.:

$$\max\left(0, g_1 + D(f(x_i),f(x_i^{p+})) - D(f(x_i),f(x_i^{p-}))\right) \le \varepsilon_i$$
$$\max\left(0, g_2 + D(f(x_i),f(x_i^{p-})) - D(f(x_i),f(x_i^{p--}))\right) \le \tau_i$$
$$\max\left(0, g_3 + D(f(x_i),f(x_i^{p--})) - D(f(x_i),f(x_i^{n}))\right) \le \sigma_i$$
$$\forall i,\ \varepsilon_i \ge 0,\ \tau_i \ge 0,\ \sigma_i \ge 0$$

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



- Training step



Every 5000 iterations

Feature-based clustering

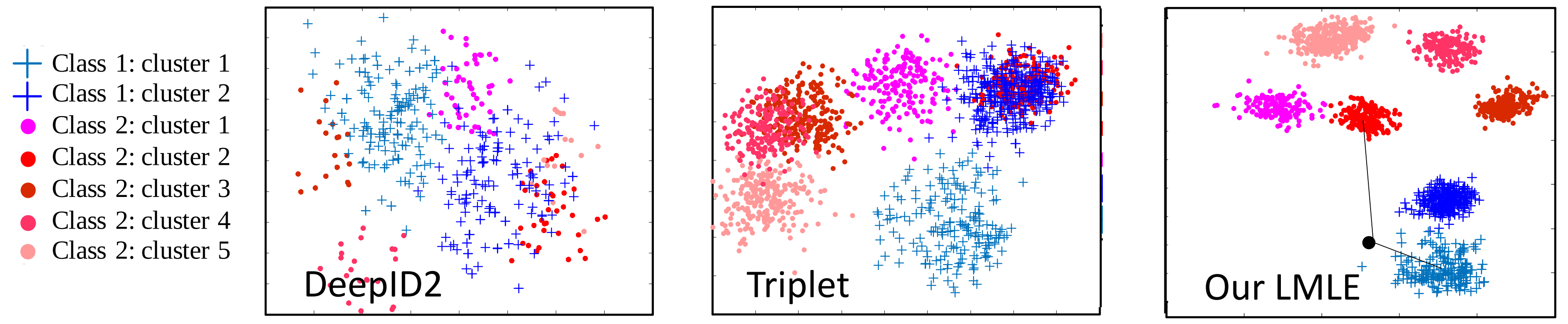
- Clustering by k-means
- Generate quintuplets from cluster & class membership

Feature learning/updating

- Re-sample batches equally from each class
- Forward their quintuplets to CNN to compute loss
- Back-propagation

4. Cluster-wise kNN search

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



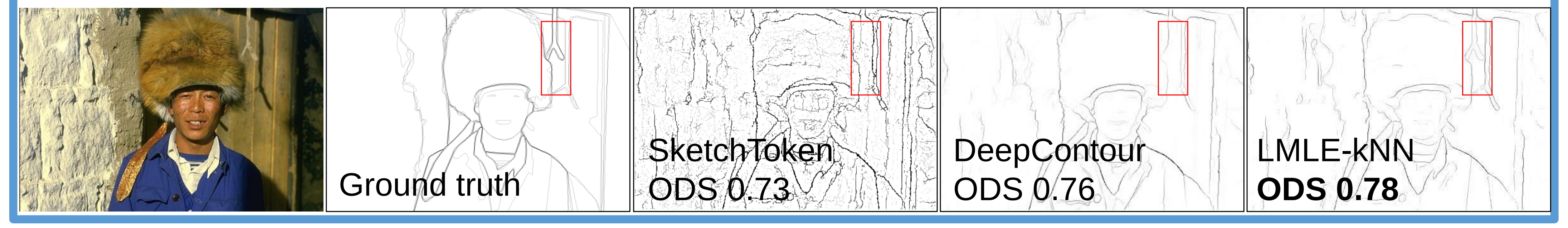
$$y_q = \underset{c=1,\dots,C}{\arg\max}\left(\min_{\substack{m_j\in\phi(q)\\ y_j\neq c}} D(f(q),f(m_j)) - \max_{\substack{m_i\in\phi(q)\\ y_i=c}} D(f(q),f(m_i))\right)$$

5. Results

- Large-scale CelebA face attribute dataset
  - **200K** celebrity images, each with **40** attributes
  - Highly imbalanced: average positive class rate **23%**
  - Total accuracy $=\left(\frac{tp+tn}{Np+Nn}\right)$ → Balanced accuracy $=\frac{1}{2}\left(\frac{tp}{Np}+\frac{tn}{Nn}\right)$

	Total acc.	Balanced acc.
Triplet-kNN	83	72
Anet	87	80
LMLE-kNN	**90**	**84**

- 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**