A tutorial on Metric Learning with Neural Network

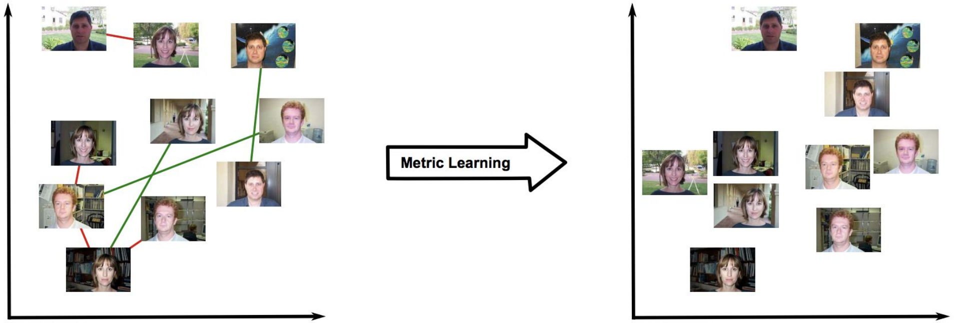
-From Siamese to Triplet

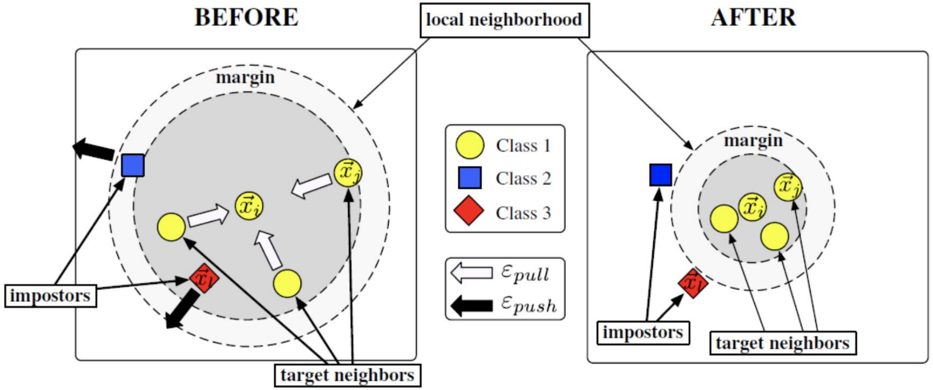
michael.lee@kakaocorp.com

Metric

- A Metric is a function that quantifies a “distance” between every pair of elements in a set, thus inducing a measure of similarity.
- Properties
  - Non-negativity
  - Identity of Discernible, f(x, y) = 0 ⇔ x = y
  - Symmetry
  - Triangle Inequality

Metric Learning



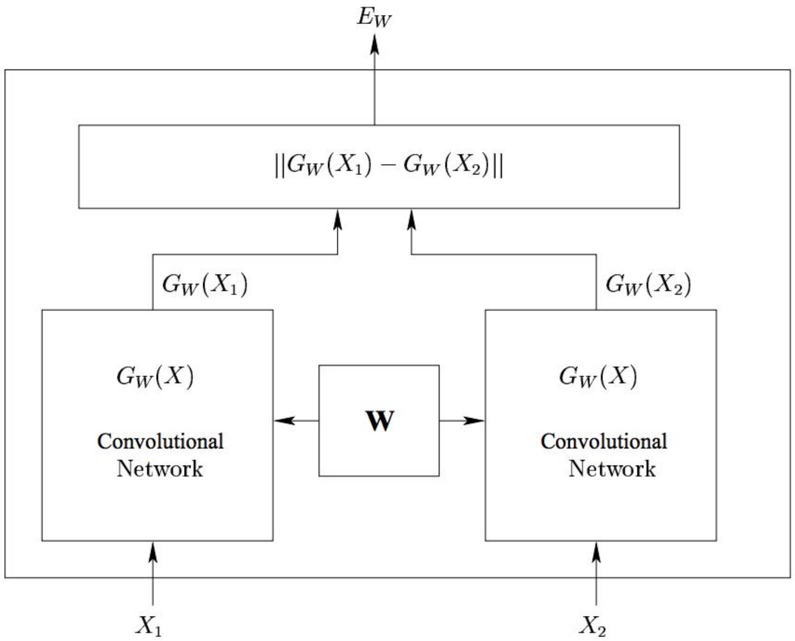


$$\min_{A\succeq 0} \sum_{(i,j)\in\mathcal{S}} d_A(\boldsymbol{x}_i,\boldsymbol{x}_j) + \lambda \sum_{(i,j,k)\in\mathcal{R}} [1 + d_A(\boldsymbol{x}_i,\boldsymbol{x}_j) - d_A(\boldsymbol{x}_i,\boldsymbol{x}_k)]_+.$$

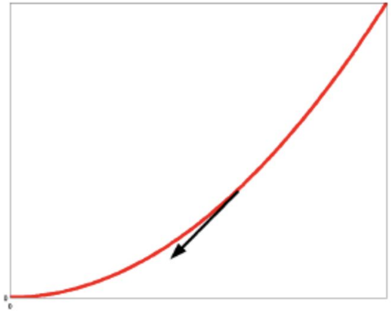
Metric Learning with Neural Networks

- Siamese Neural Networks
- Triplet Neural Networks
    - NCA Loss, MagNet, etc.

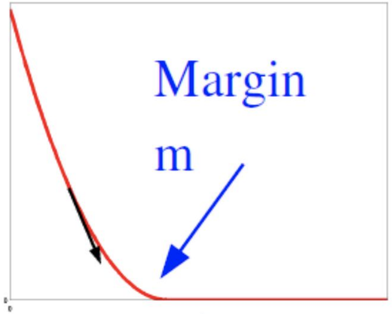
Siamese Network



$$L_{similar} = \frac{1}{2} D_w^2$$



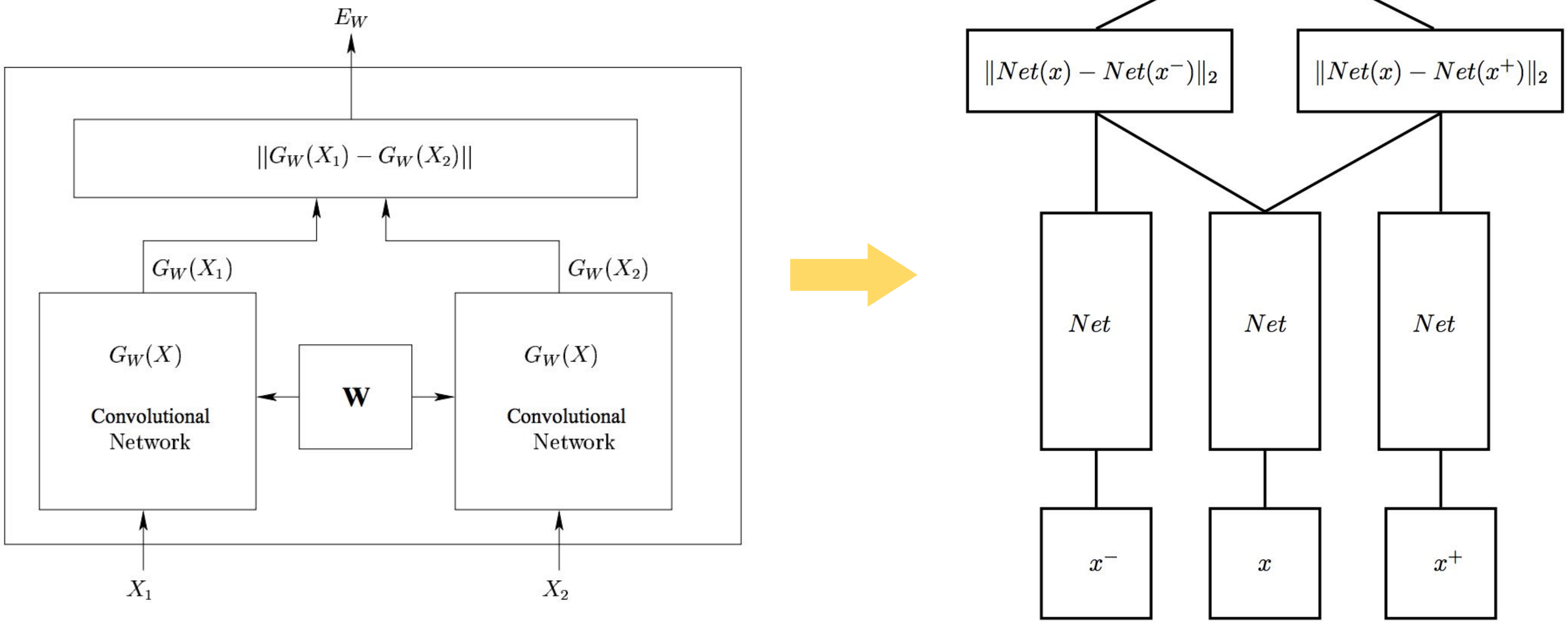
$$L_{dissimilar} = \frac{1}{2} \{max(0, m - D_W)\}^2$$



$$L(W, Y, \vec{X_1}, \vec{X_2}) =$$

$$(1 - Y)\frac{1}{2}(D_W)^2 + (Y)\frac{1}{2}\{max(0, m - D_W)\}^2$$

Triplet Network



Loss Function for Triplet Network

$$Loss(d_+, d_-) = \|(d_+, d_- - 1)\|_2^2 = const \cdot d_+^2$$

$$d_+ = \frac{e^{\|Net(x) - Net(x^+)\|_2}}{e^{\|Net(x) - Net(x^+)\|_2} + e^{\|Net(x) - Net(x^-)\|_2}}$$

$$d_- = \frac{e^{\|Net(x) - Net(x^-)\|_2}}{e^{\|Net(x) - Net(x^+)\|_2} + e^{\|Net(x) - Net(x^-)\|_2}}$$

Application 1 - Face Recognition and Clustering

Loss function

$$\sum_{i}^{N}\left[\|f(x_i^a)-f(x_i^p)\|_2^2-\|f(x_i^a)-f(x_i^n)\|_2^2+\alpha\right]_+$$

Semi hard examples sampling

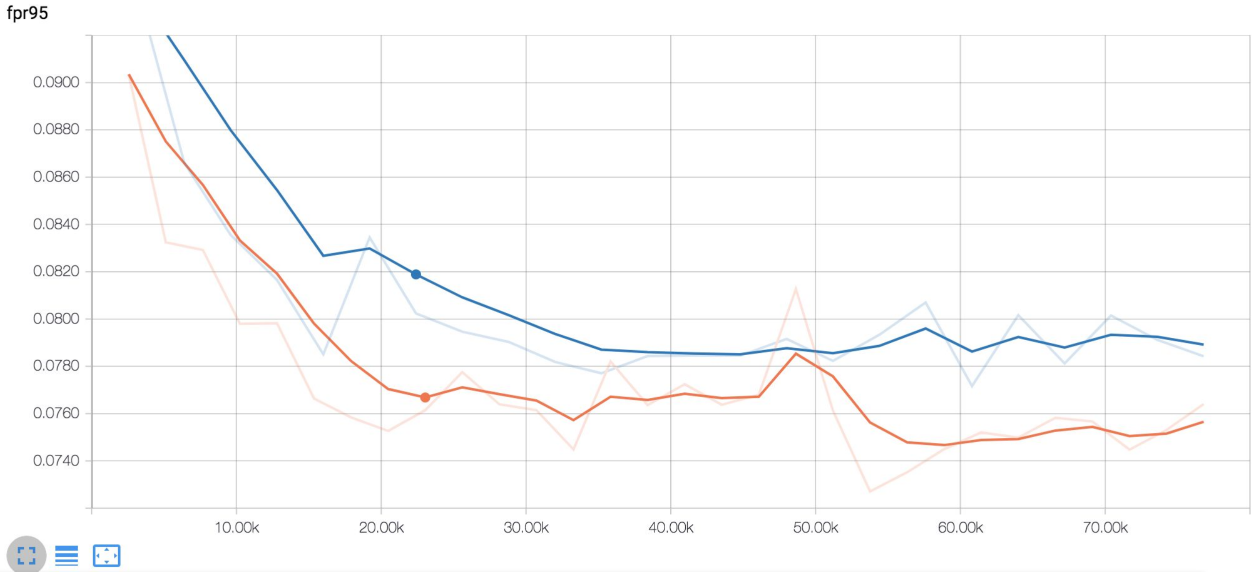
$$\|f(x_i^a)-f(x_i^p)\|_2^2<\|f(x_i^a)-f(x_i^n)\|_2^2$$

Application 1 - Cont’d

- Code: https://github.daumkakao.com/michael-lee/pytorch-triplet
- Test dataset: Tiny ImageNet
- Classes: 200
- Validation set: 128000 pair
- Embedding Size: 128 dim
- Measurement: FPR95 (False Positive Rate at True Positive Rate 95%)

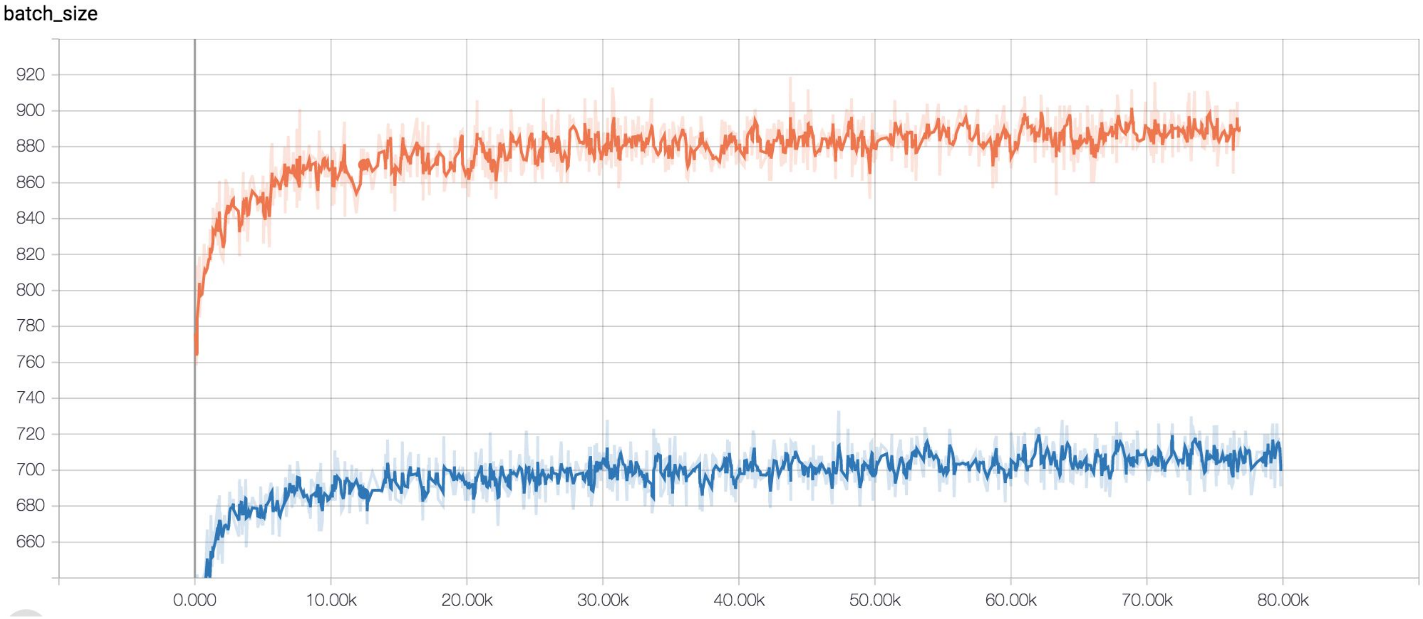
Application 1 - Cont’d

Performance comparison (random sample vs semi hard example sampling)

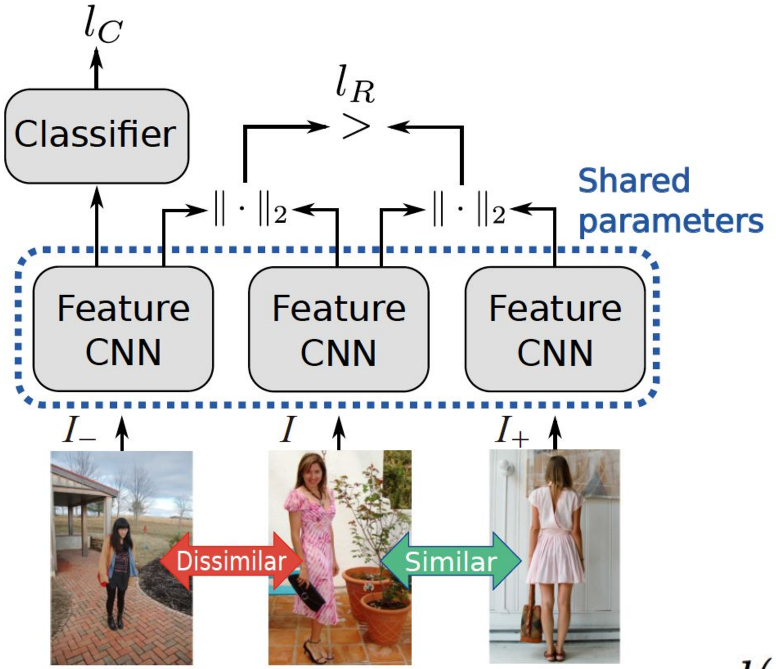


Application 1 - Cont’d

Mini-batch size after mining



Application 2 - Triplet Network with Classification



$$d_- = \frac{\exp(\|\boldsymbol{f}_- - \boldsymbol{f}\|_2)}{\exp(\|\boldsymbol{f}_- - \boldsymbol{f}\|_2) + \exp(\|\boldsymbol{f}_+ - \boldsymbol{f}\|_2)}$$

$$d_+ = \frac{\exp(\|\boldsymbol{f}_+ - \boldsymbol{f}\|_2)}{\exp(\|\boldsymbol{f}_- - \boldsymbol{f}\|_2) + \exp(\|\boldsymbol{f}_+ - \boldsymbol{f}\|_2)}$$

$$l_R(d_+, d_-) = 0.5\left((d_+)^2 + (1 - d_-)^2\right) = (d_+)^2$$

Multi label Cross entropy

$$l_C(X_-, \boldsymbol{y}_-) = \frac{1}{|T|} \sum_{t \in T} l_\times(X_-^t,\ y_-^t)\ ,$$

$$l_\times(x, y) = -x_y + \log\left(\exp(x_0) + \exp(x_1)\right)$$

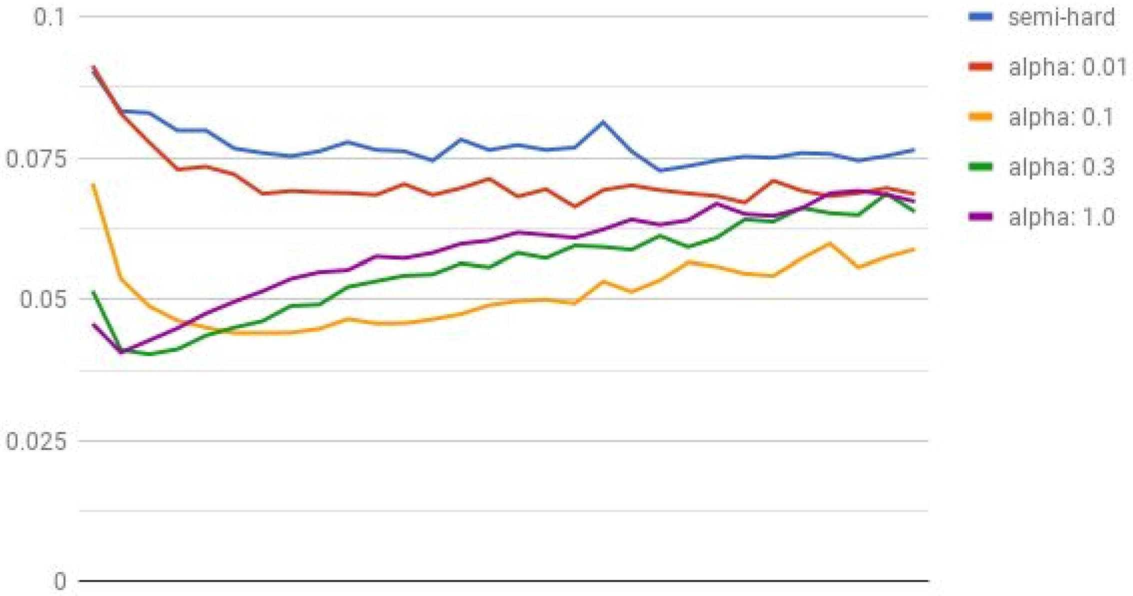
$$l(d_+, d_-, X_-, \boldsymbol{y}_-) = l_R(d_+, d_-) + \alpha l_C(X_-, \boldsymbol{y}_-)$$

Application 2 - Cont’d

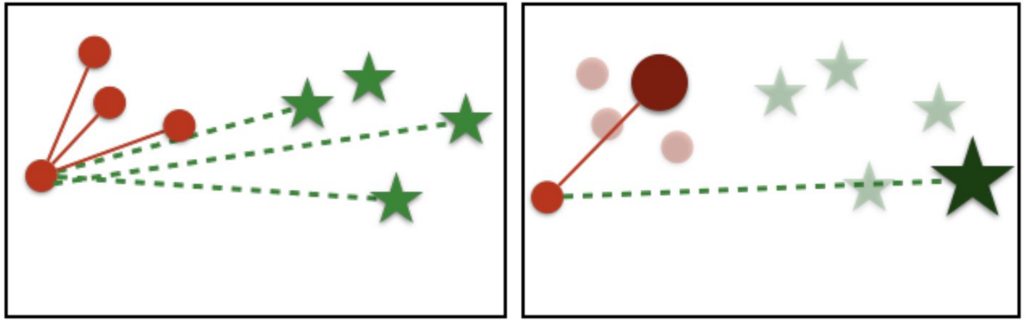
- Code: https://github.daumkakao.com/michael-lee/pytorch-triplet-cls
    - Differences
        - Single Label Classification
        - Classification Networks for the anchor
        - Data Sampling
- Test dataset: Tiny ImageNet
- Classes: 200
- Validation set: 128000 pair
- Embedding Size: 128 dim
- Measurement: FPR95 (False Positive Rate at True Positive Rate 95%)

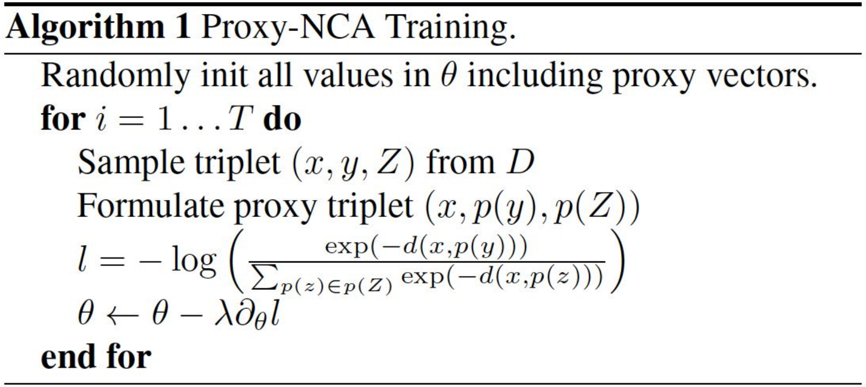
Application 2 - Cont’d

Performance comparison (Triplet vs Triplet with Classification)



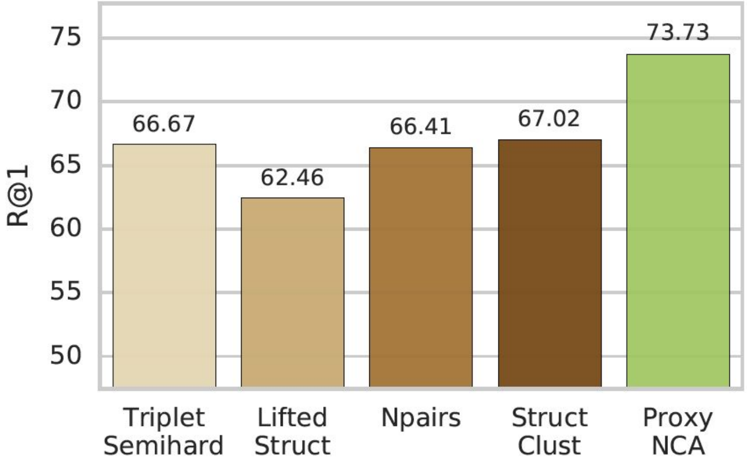
Application 3 - Proxy Ranking Loss





Algorithm 1 Proxy-NCA Training.

```
Randomly init all values in $\theta$ including proxy vectors.
for $i = 1 \ldots T$ do
    Sample triplet $(x, y, Z)$ from $D$
    Formulate proxy triplet $(x, p(y), p(Z))$
    $l = -\log\left(\frac{\exp(-d(x,p(y)))}{\sum_{p(z)\in p(Z)} \exp(-d(x,p(z)))}\right)$
    $\theta \leftarrow \theta - \lambda \partial_\theta l$
end for
```



References

1. Learning a similarity metric discriminatively, with application to face verification, CVPR 2005
2. A Survey on Metric Learning for Feature Vectors and Structured Data, arXiv 2013
3. Siamese Neural Networks for One-shot Image Recognition, 2015
4. Metric Learning: A Survey, Foundations and Trends® in Machine Learning, 2013
5. Deep Metric Learning Using Triplet Network, ICLR Workshop 2015
6. FaceNet: A Unified Embedding for Face Recognition and Clustering, CVPR 2015
7. Fashion Style in 128 Floats: Joint Ranking and Classification using Weak Data for Feature Extraction, CVPR 2016
8. No Fuss Distance Metric Learning using Proxies, ICCV 2017