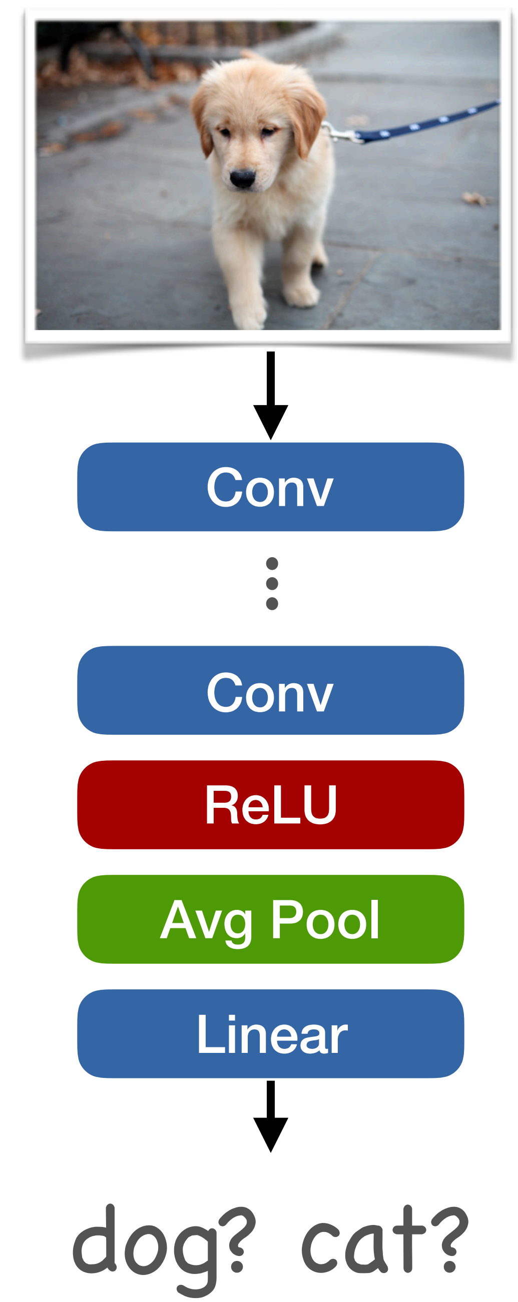
Learning with an expanding set of labels

Recap: Classification

- Collect dataset
- Label dataset
- Train your big model



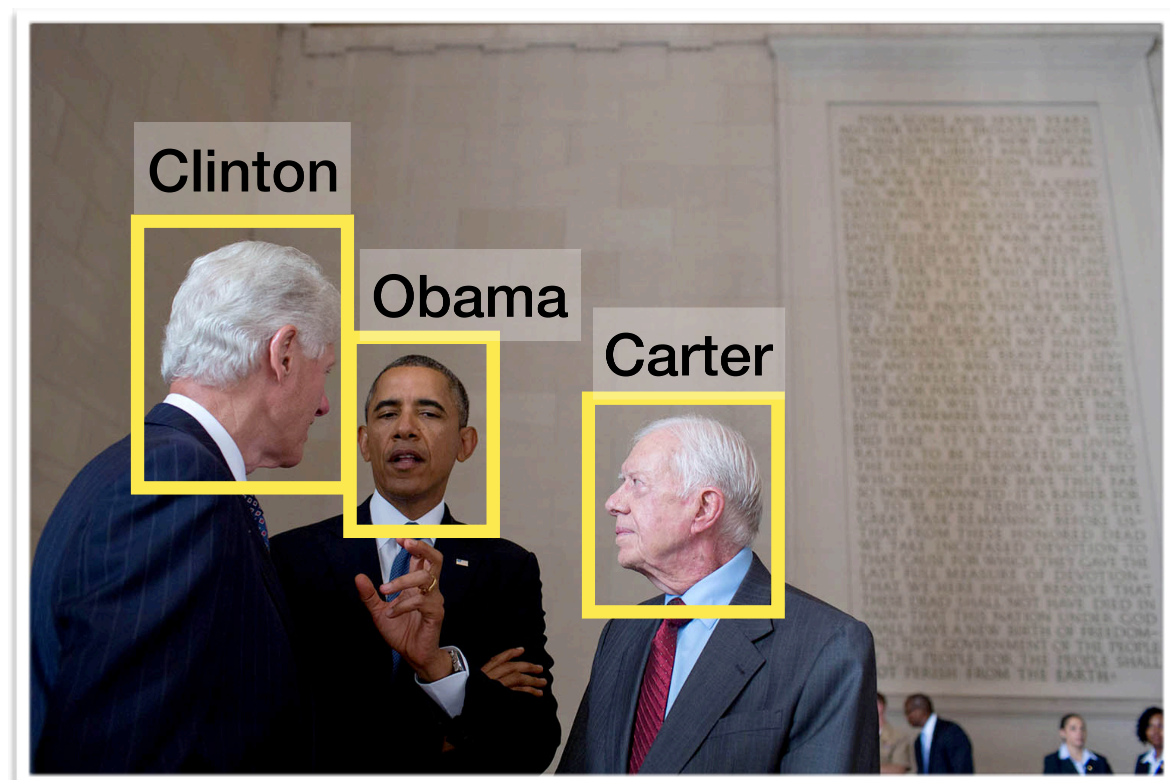
Example: Face classification


Bush


Kennedy


Washington

- Classify who is in a picture
  - Each person is a class



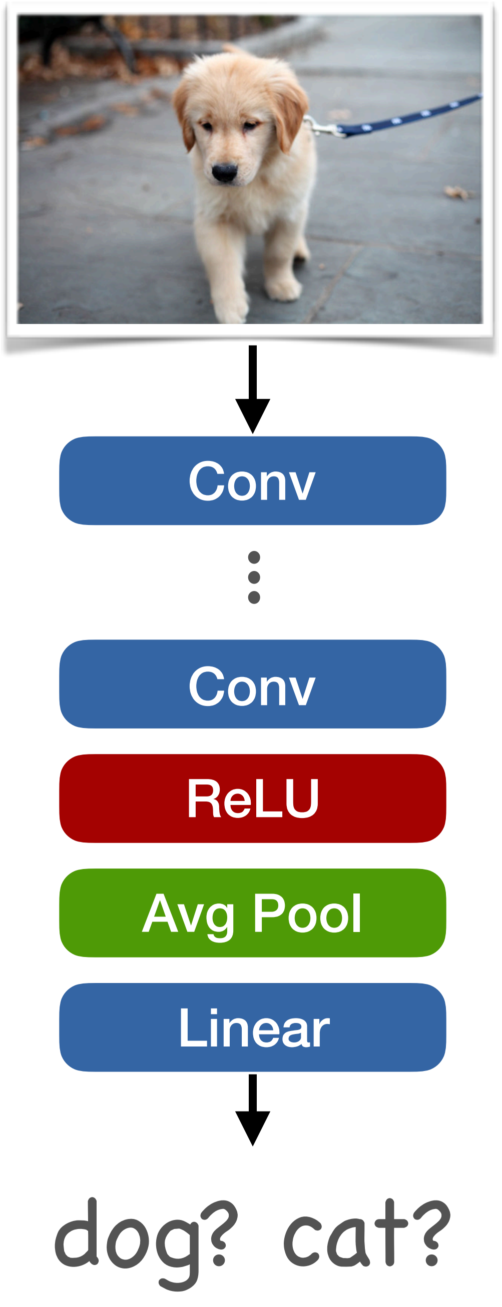
Issues

- What do we do when we have a new class?
  - Classifier needs to retrain

Embedding learning

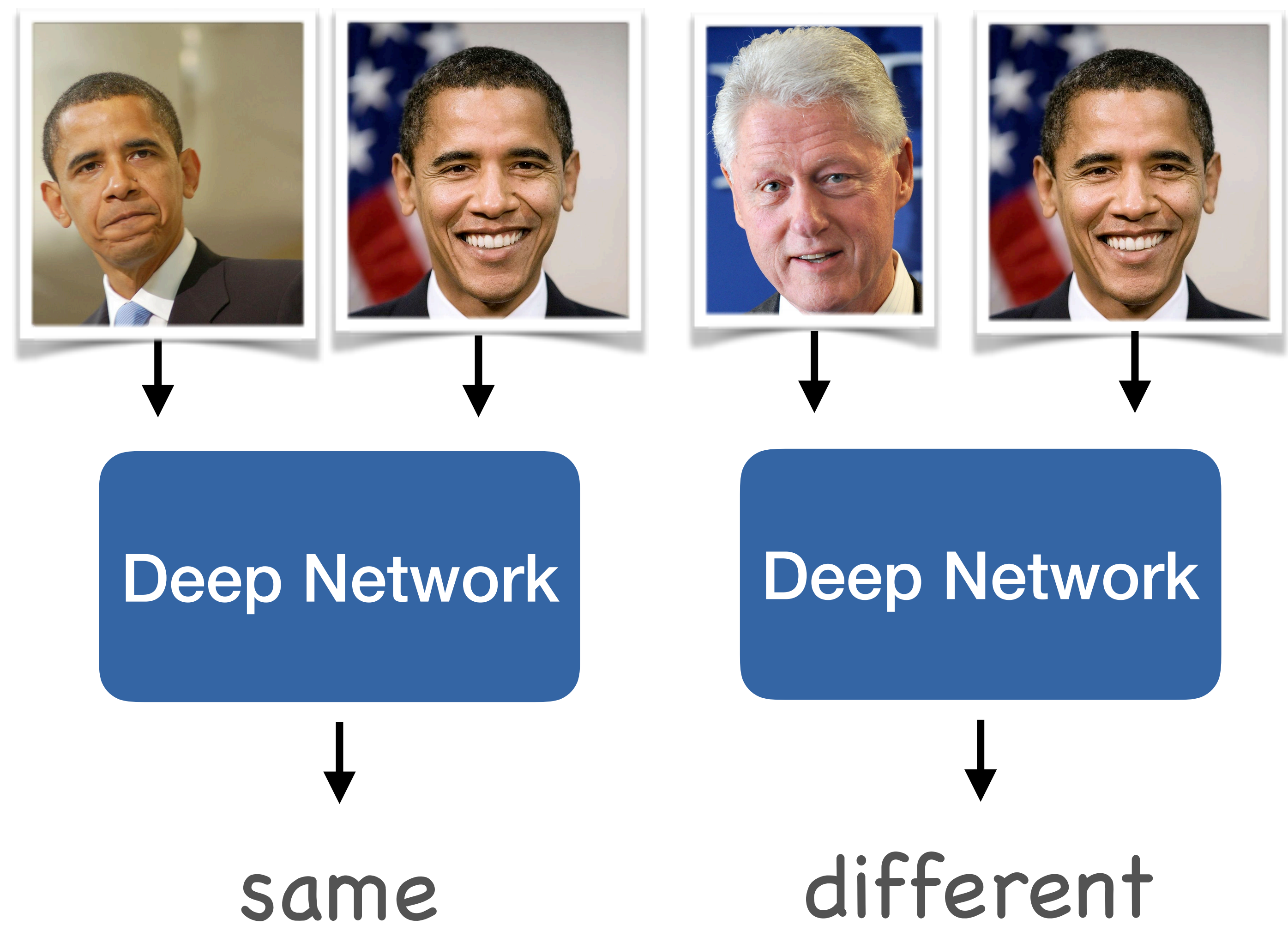
Expanding label set

- Standard classification requires fixed label set



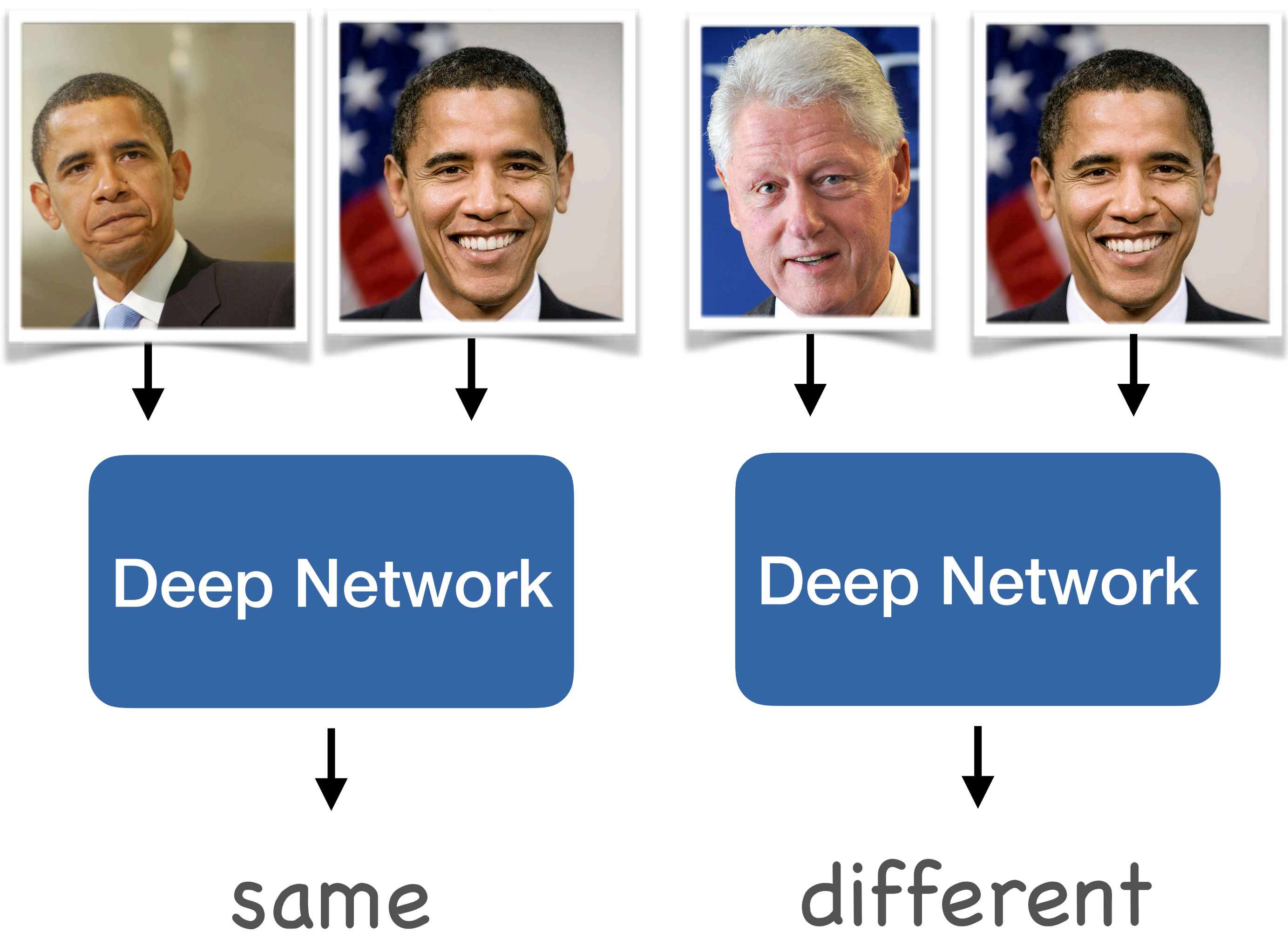
How do we avoid re-training?

- Switch classification
  - same person or not?



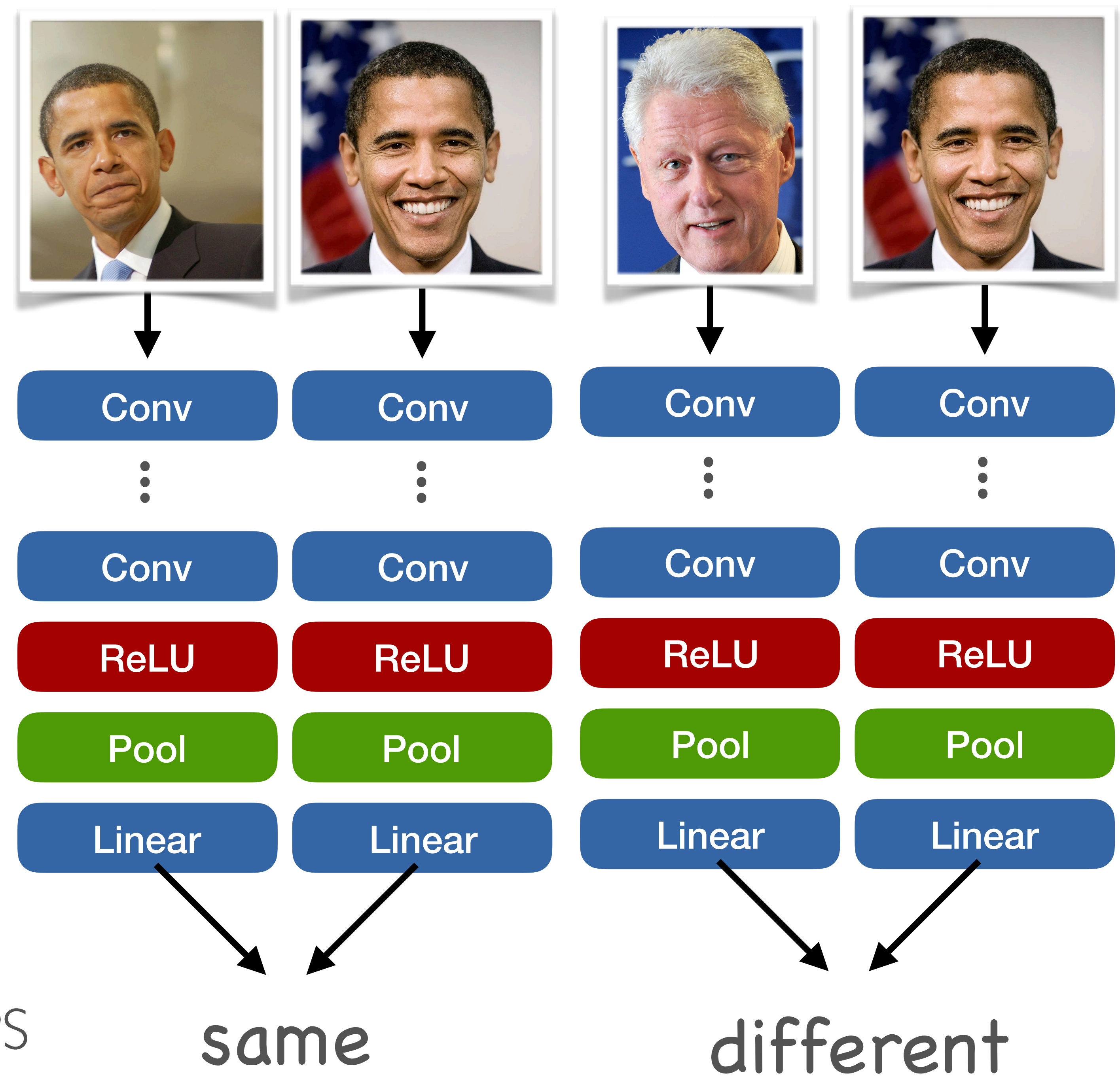
Learning similarity - issues

- Very slow inference
  - Forward pass for each pair



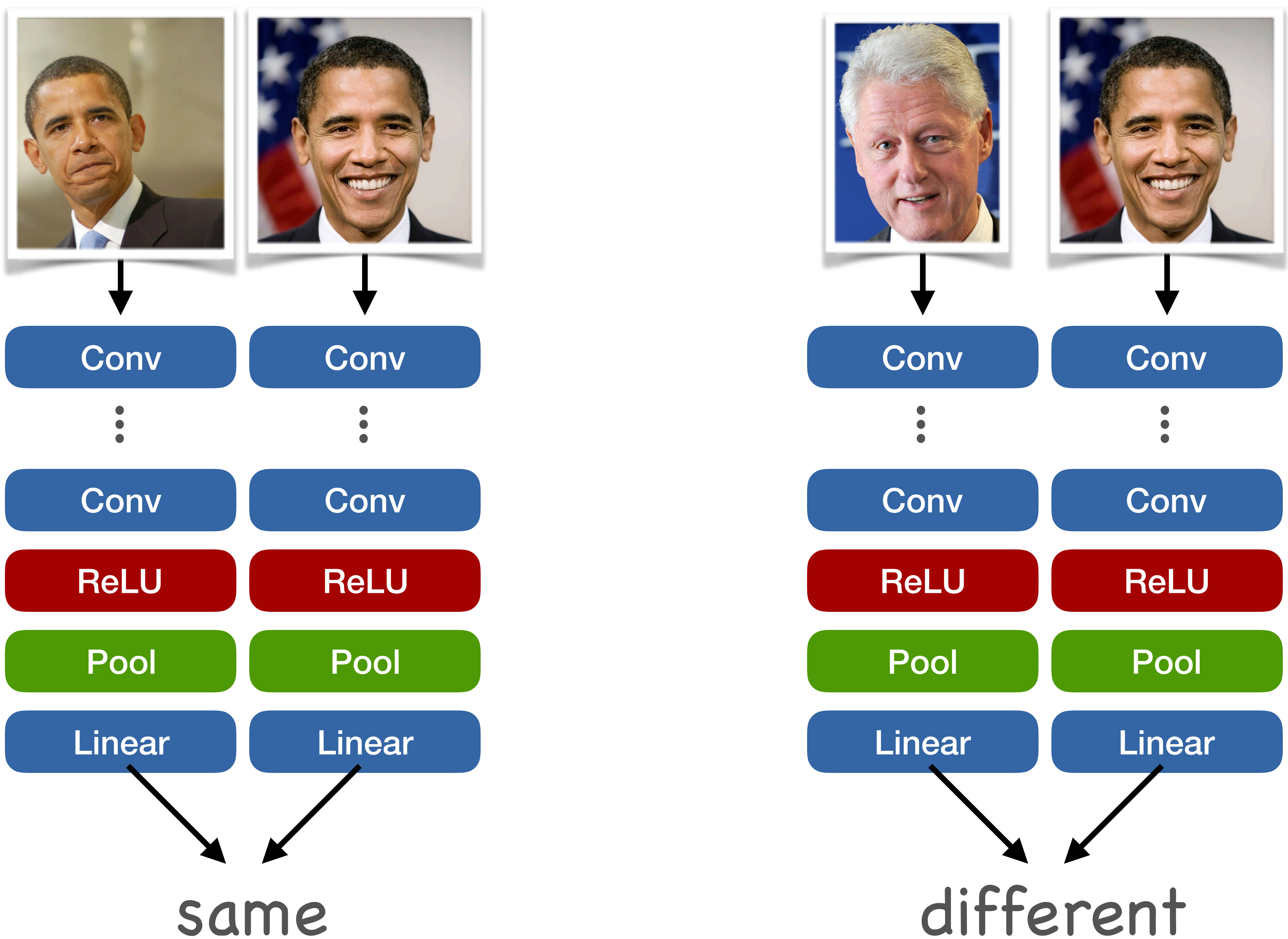
Solution - Siamese networks

- Embedding network
  - One per image
- Distance metric
  - E.g. only dot product
  - Run on all pairs
    - or KNN search



Signature Verification using a Siamese Time Delay Neural Network, Bromley et al., NIPS 1994

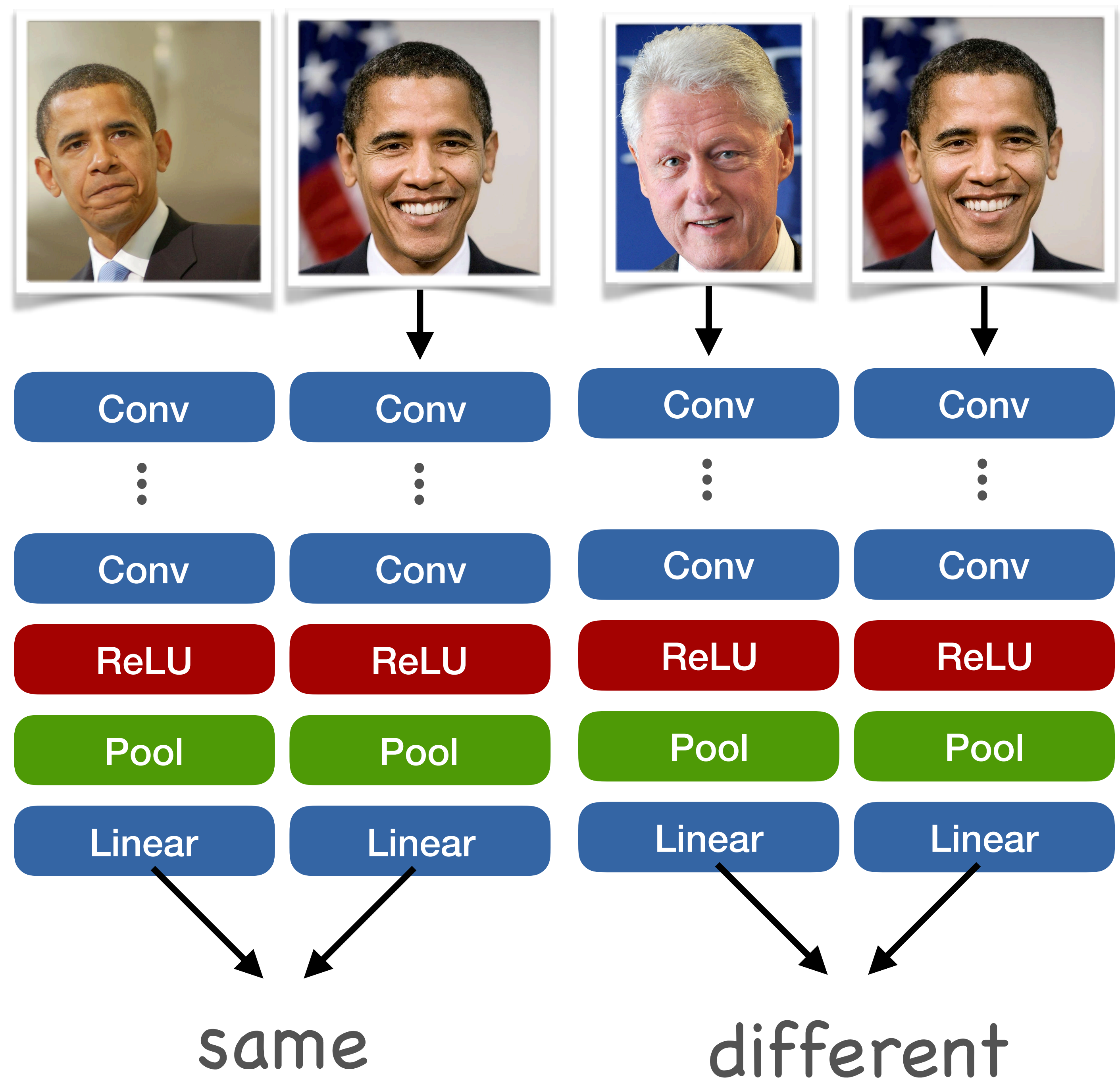
How do we train siamese networks?



Contrastive loss

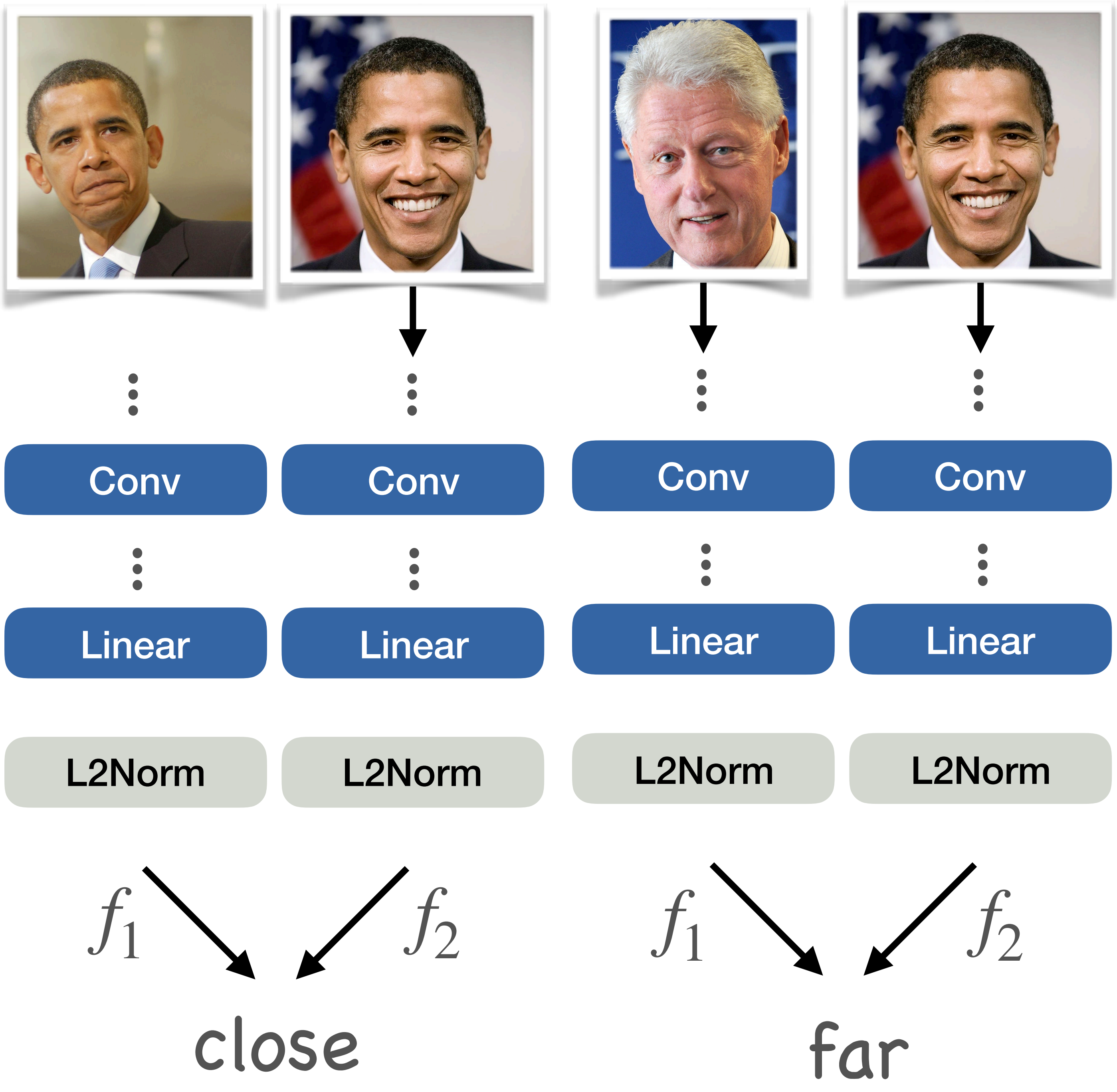
Embedding learning

- Embedding network
  - Positive and negative samples



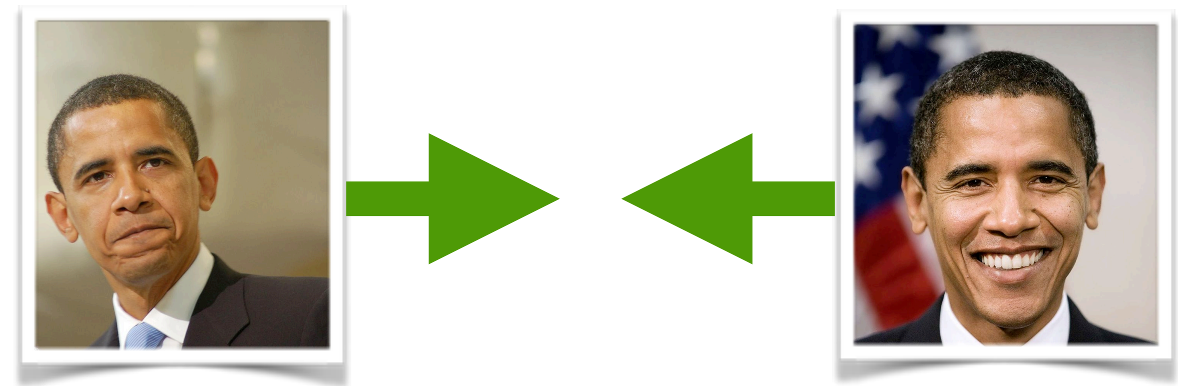
Objective

- Positives
    - $\|f_1 - f_2\| < c$
- Negatives
    - $\|f_1 - f_2\| > c$

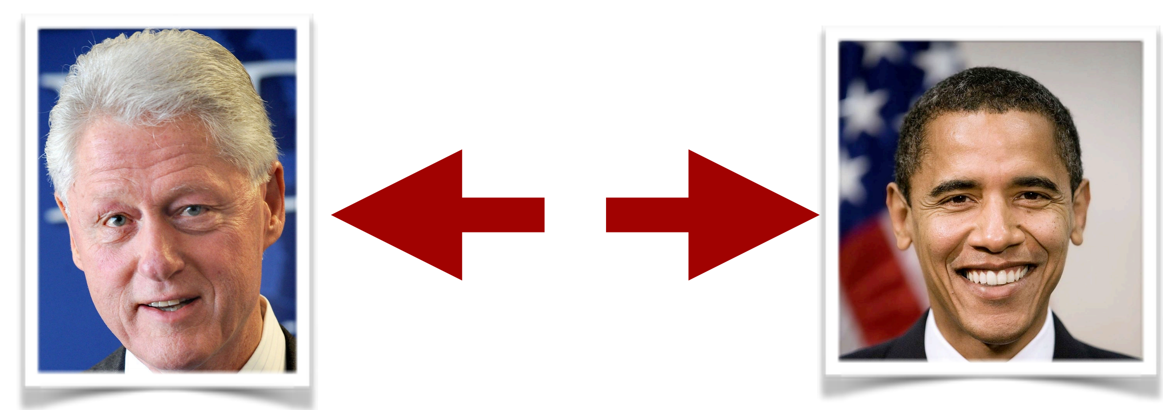


Contrastive loss

- Collapse positives
  - $||f_i - f_j||$



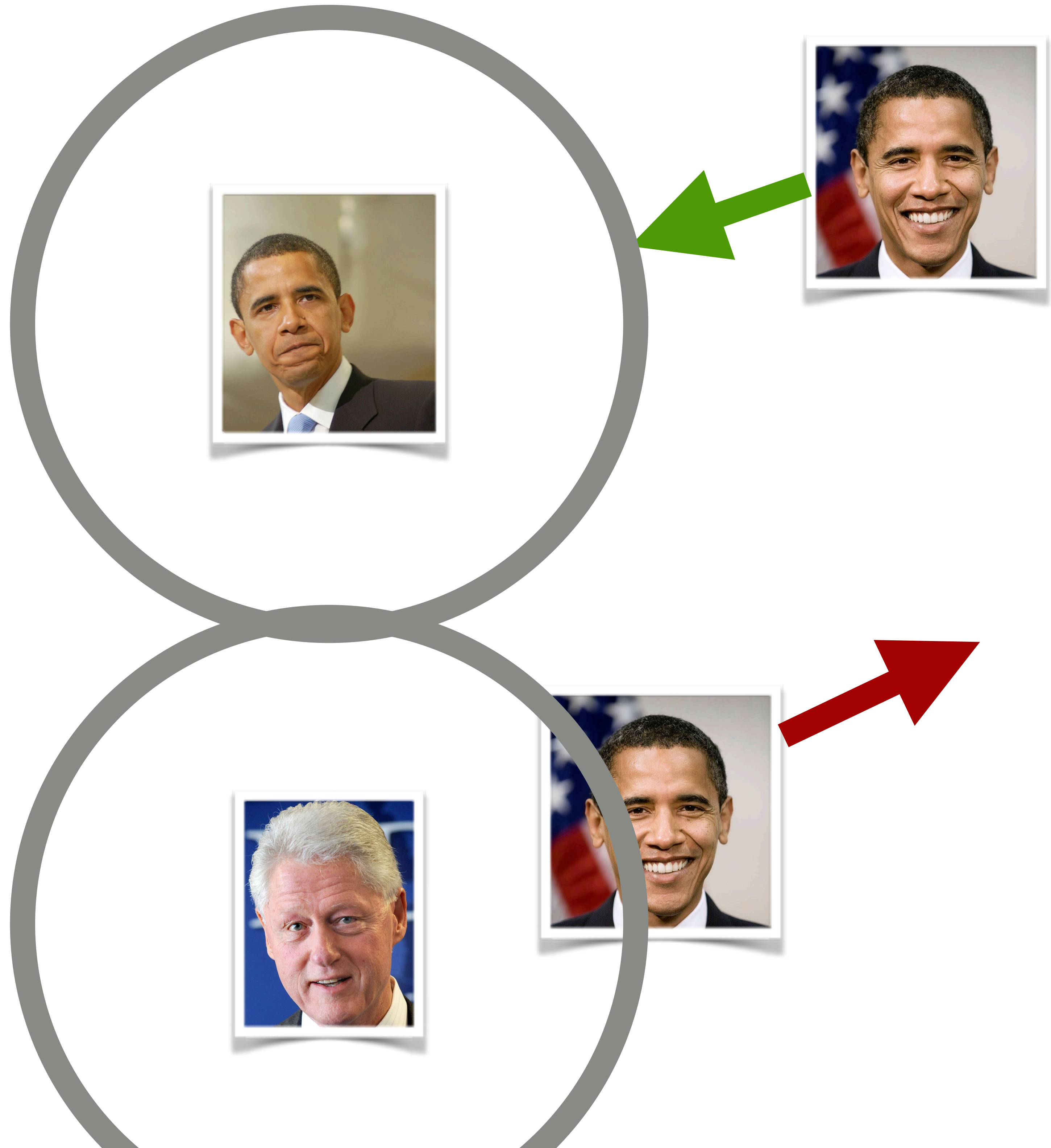
- Separate negatives
  - $\max(c - ||f_i - f_j||, 0)$



Dimensionality reduction by learning an invariant mapping, Hadsell et al., CVPR 2006

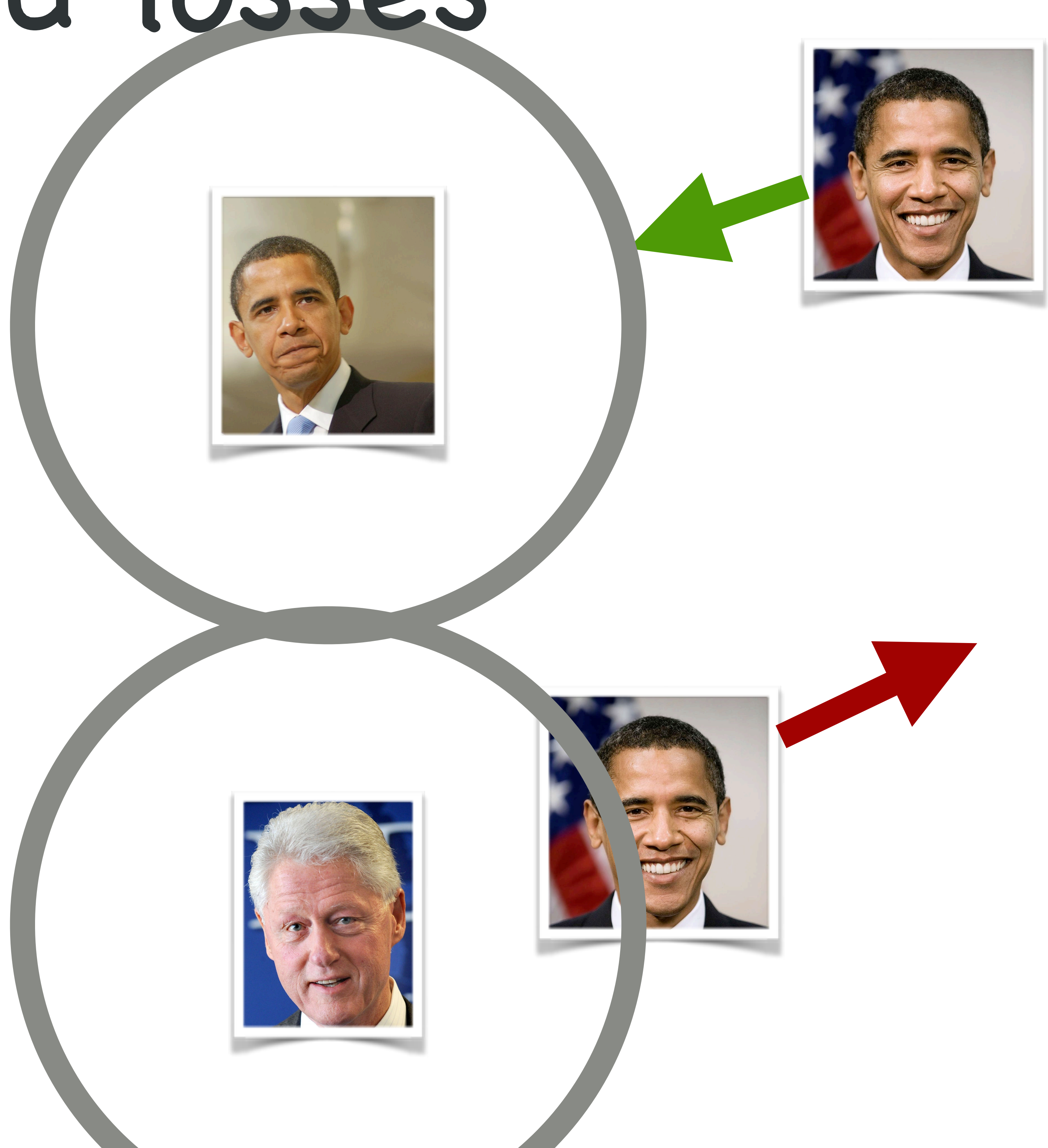
Margin based loss

- Collapse positives
  - $\max(\|f_i - f_j\| - c, 0)$
- Separate negatives
  - $\max(c - \|f_i - f_j\|, 0)$



Sampling Matters in Deep Embedding Learning, Wu et al., ICCV 2017

Contrastive and margin based losses



- Fix thresholds

Triplet loss

Embedding learning

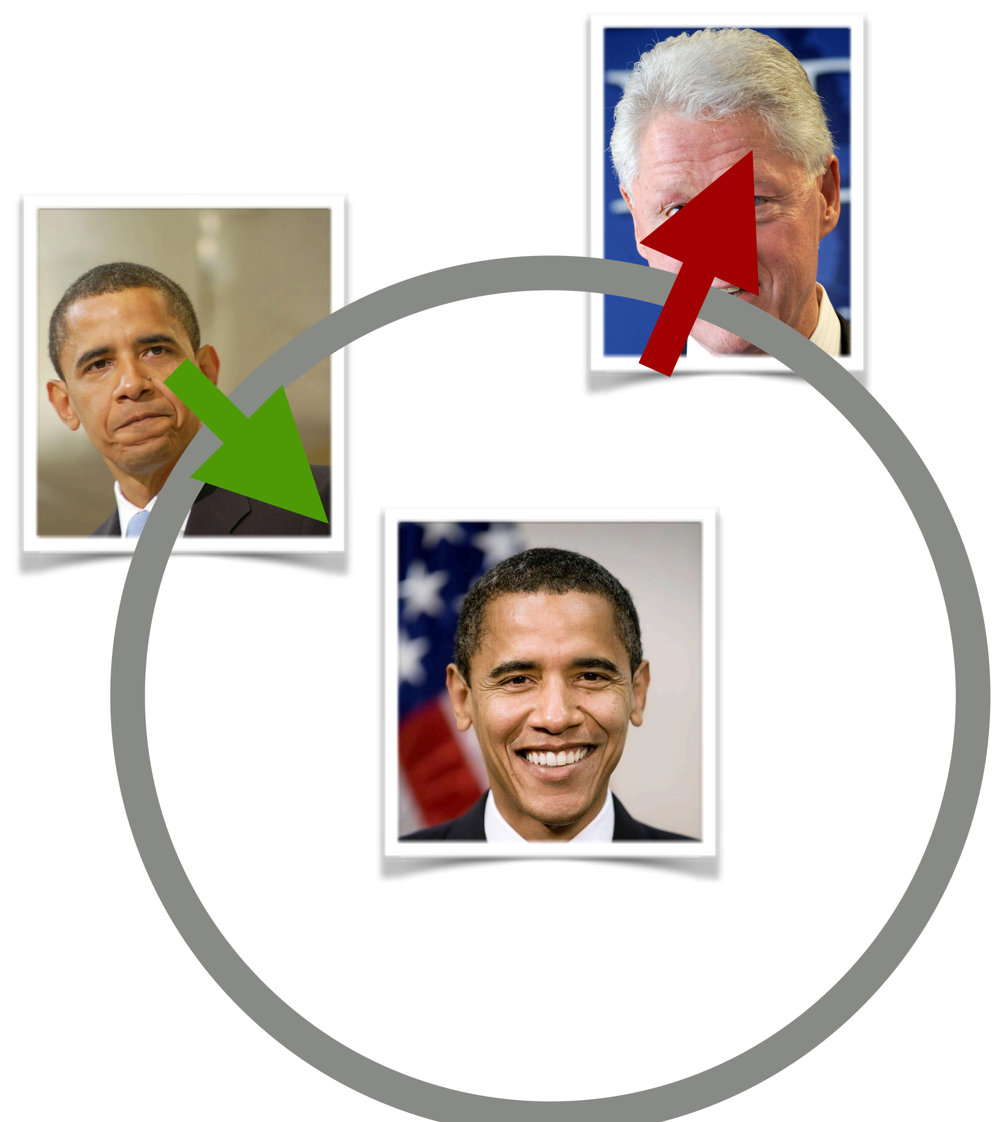
- Distances
  - Absolute distances don't matter at inference
  - KNN cares about relative distance

Triplet loss

- Objective
  $\max(0, \|f_i - f_j\| - \|f_j - f_k\| + \alpha)$

- Positive pair i, j

- Negative pair i, k



- Learning a distance metric from relative comparisons, Schultz and Joachims, NIPS 2003
- Distance metric learning for large margin nearest neighbor classification, Weinberger and Saul, JMLR 2009

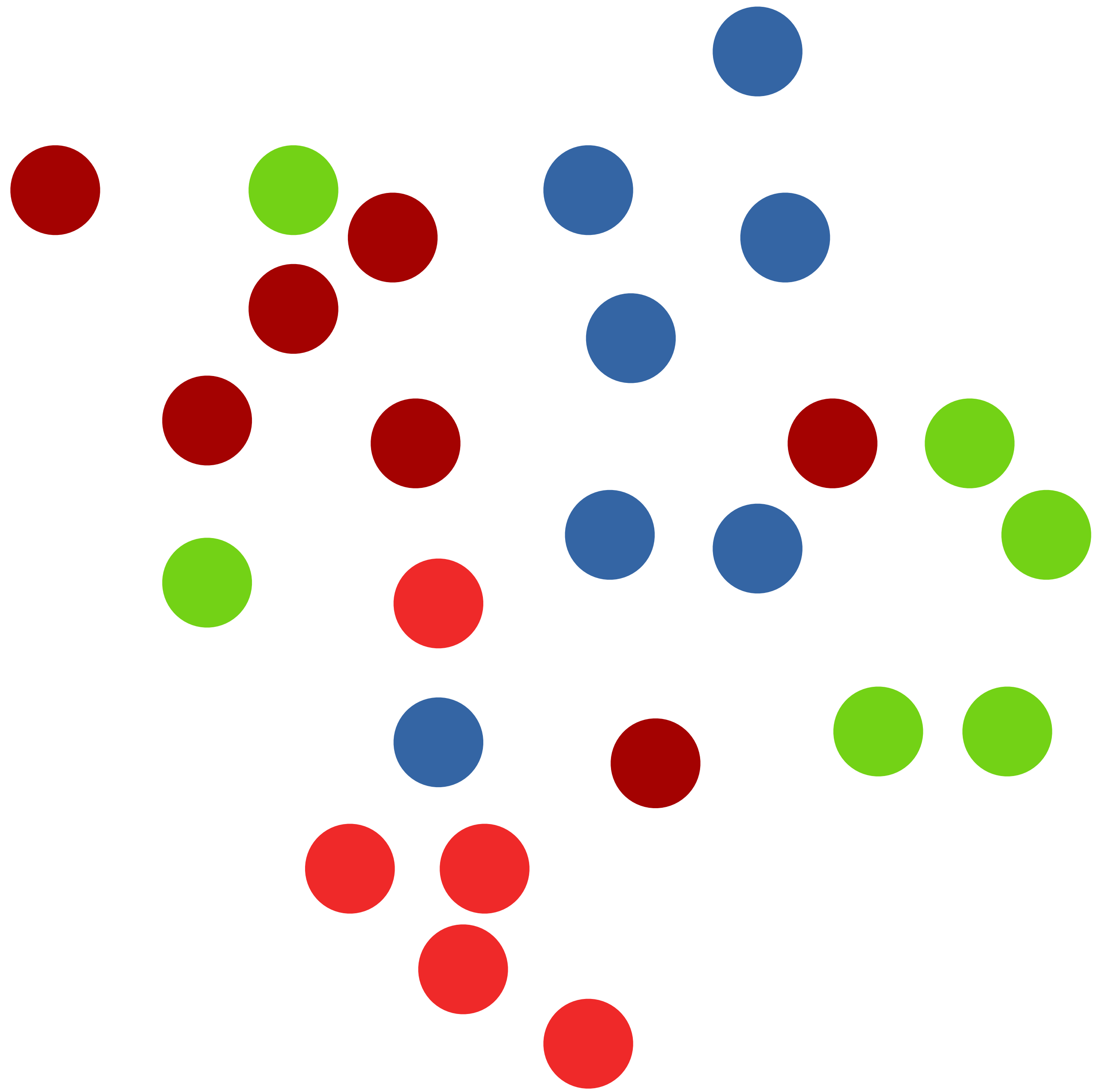
Triplet loss

- Only relative distance
  - Can deform embedding space
- Harder to train
  - All triplets vs pairs

Selecting training examples

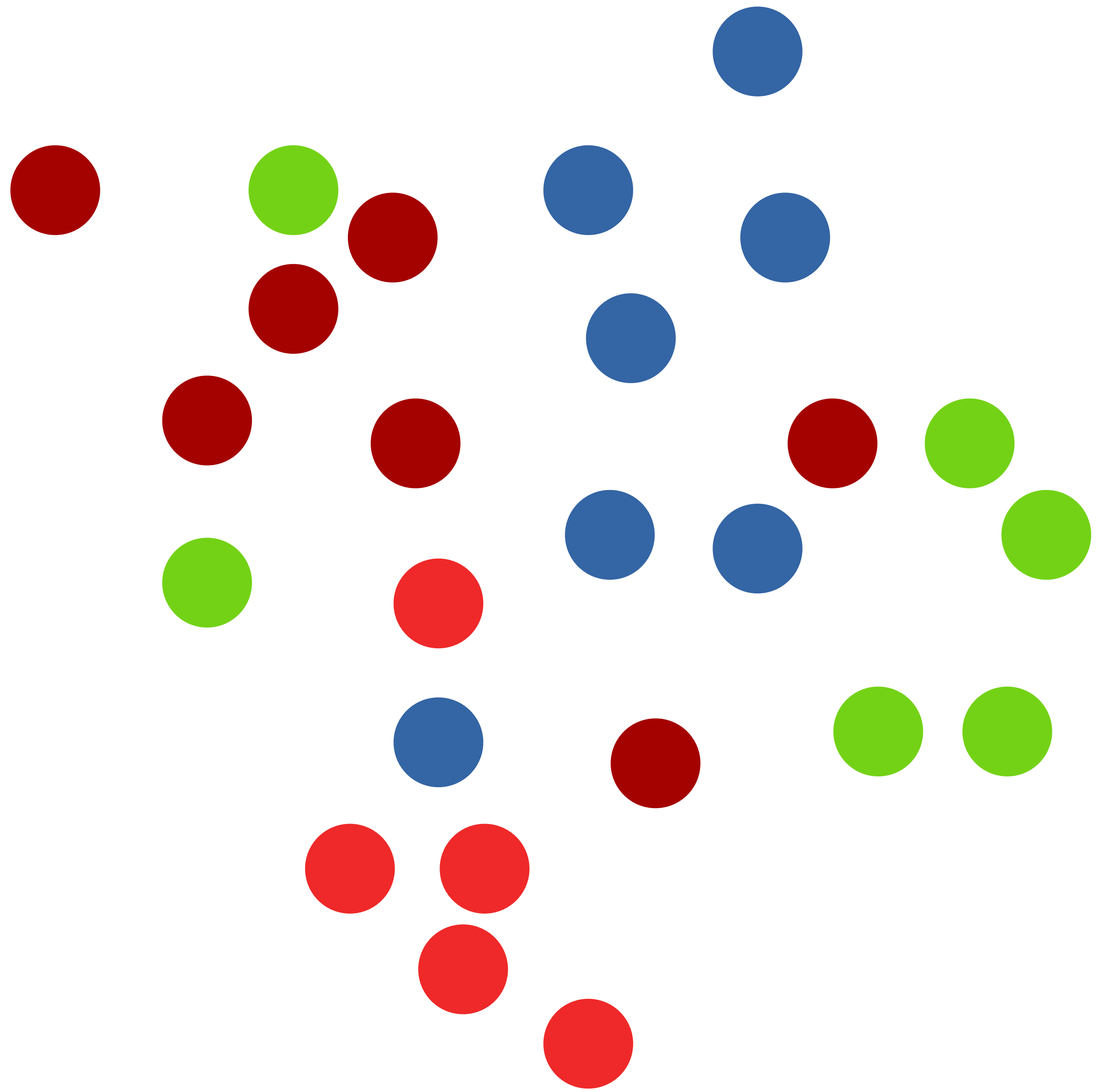
Sampling

- How do we select positives and negatives?



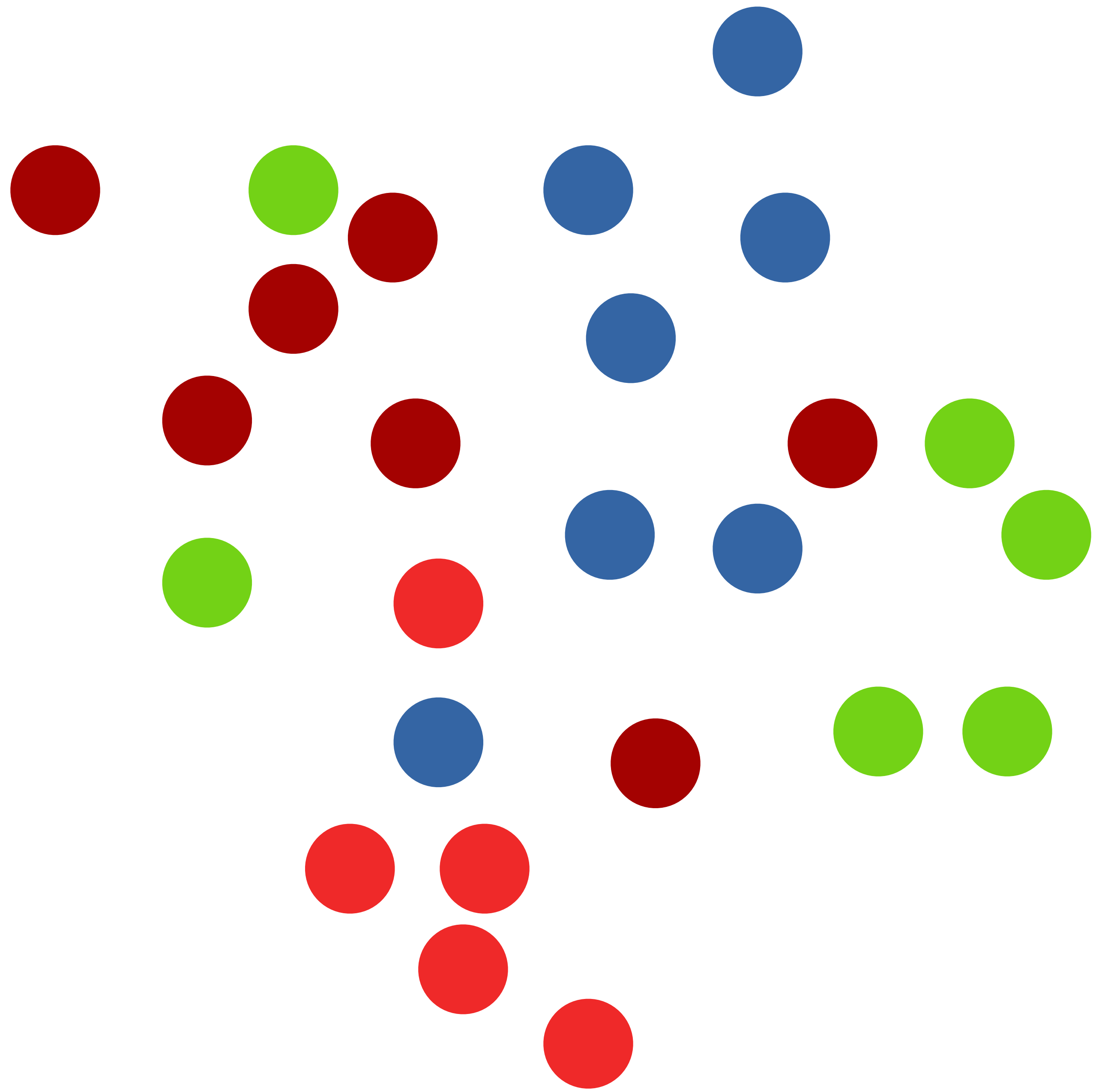
All pairs / triples?

- Bad idea
- very slow
    - Pairs $O(N^2)$
    - Triples $O(N^3)$



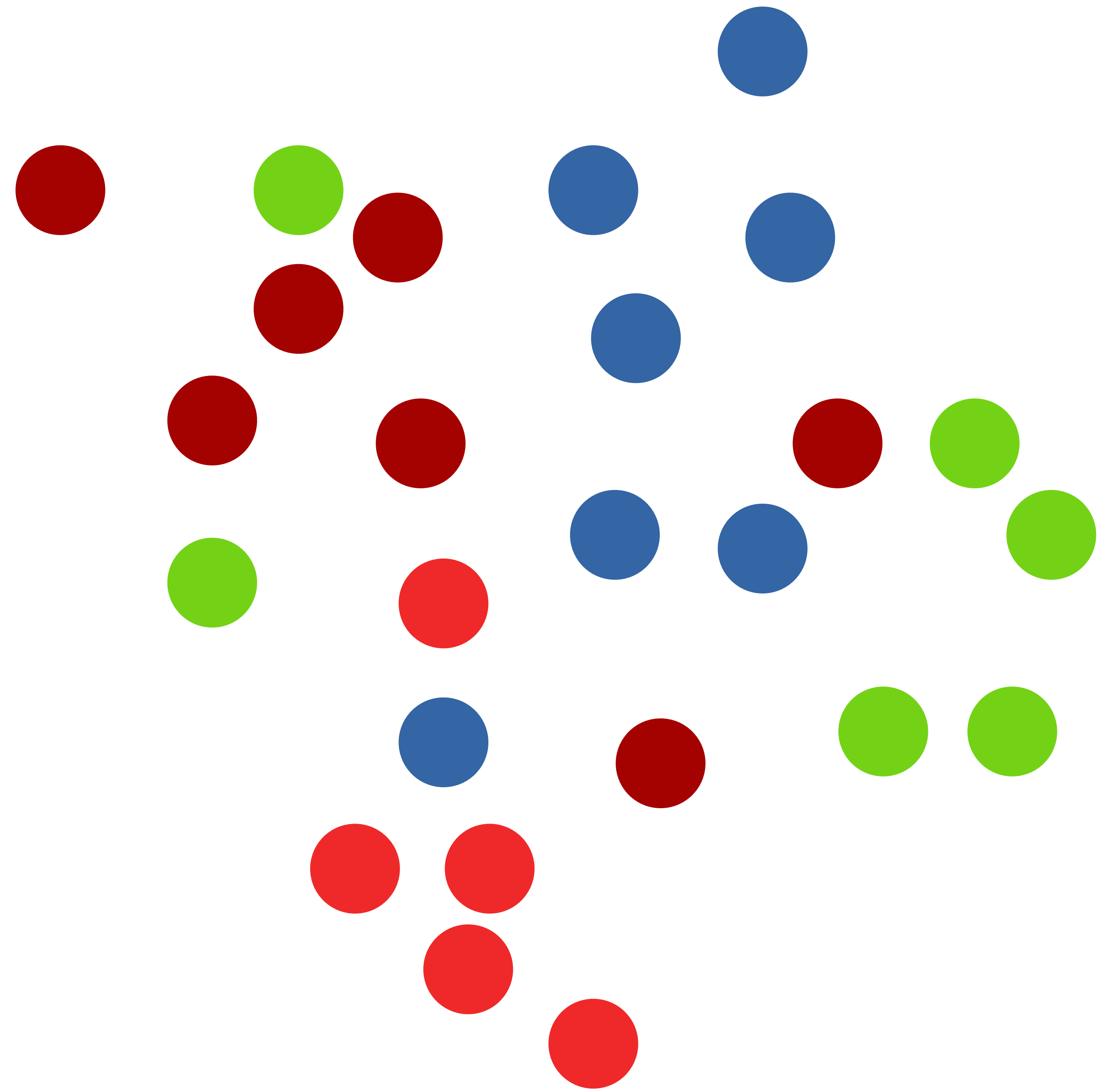
Random pairs / triples?

- Random positives
  - Fast
  - Good gradient
- Random negatives
  - Far apart
  - Small loss
  - Small gradient



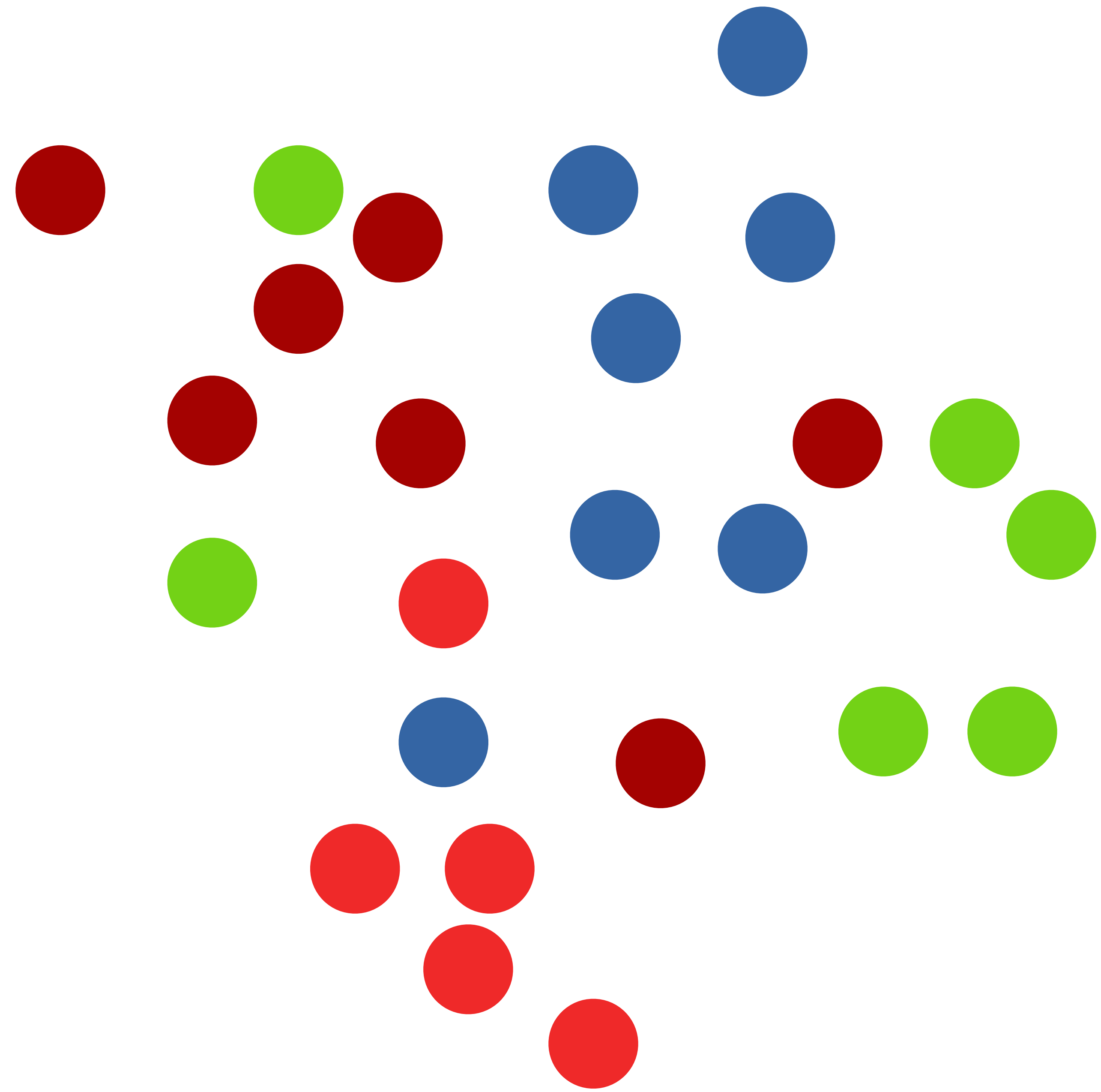
Hard negatives

- Pick one negative
  - Closed to each positive



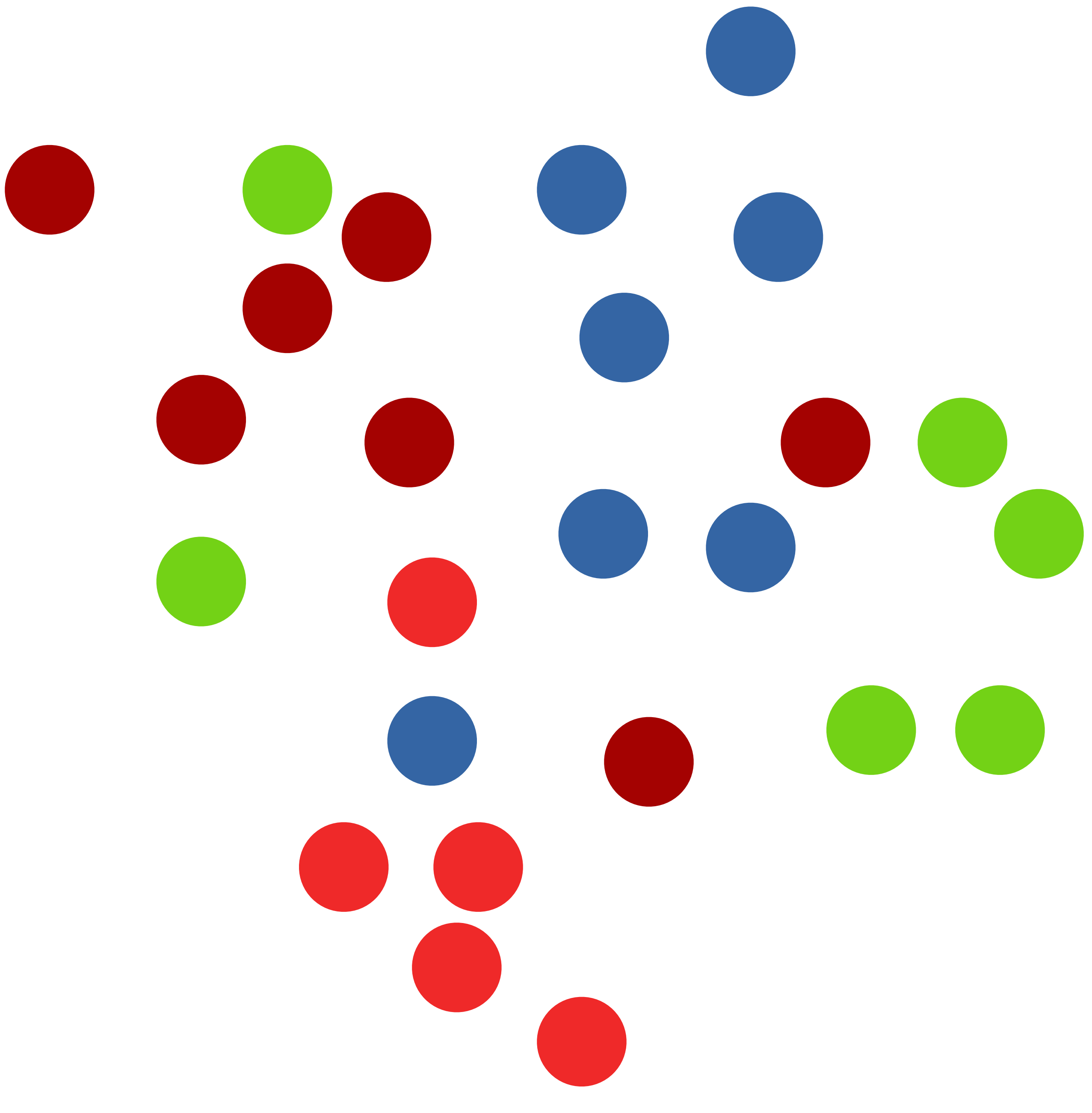
Hard negatives

- Too noisy
  - No meaningful gradient direction
- Too hard
  - Stronger gradient than positives



Semi-hard negatives

- Fine one negatives
  - at same distance as a positive



Semi-hard negatives

- Works well enough in practice
- A bit hacky
- Alternative: Weighted random sampling

Summary

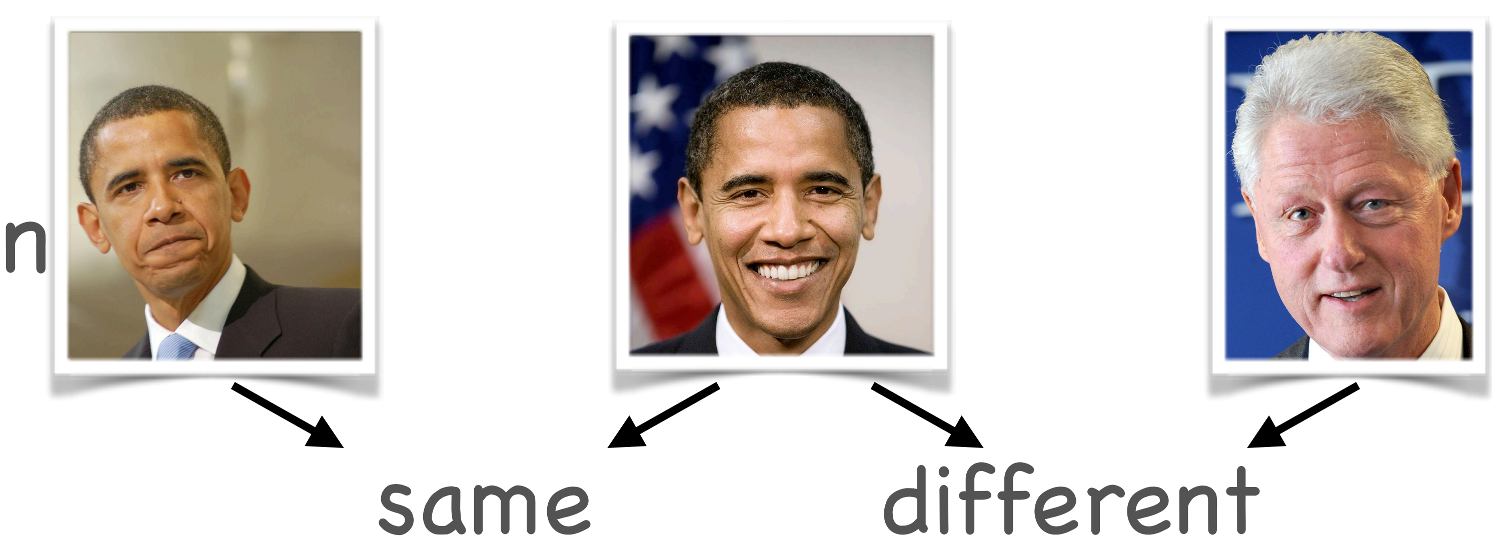
Embedding learning

- Learns with ever expanding label set
- Used in many large scale DL applications

Example: Face recognition

- Fixed label set
  - Face detection

- Expanding label set
  - Face re-identification



Example: Image retrieval and visual search

- No fixed label set
  - Sometimes: No labels at all
  - Comparisons only

query

results