Lecture 11: Deep Metric Learning

Metric Learning

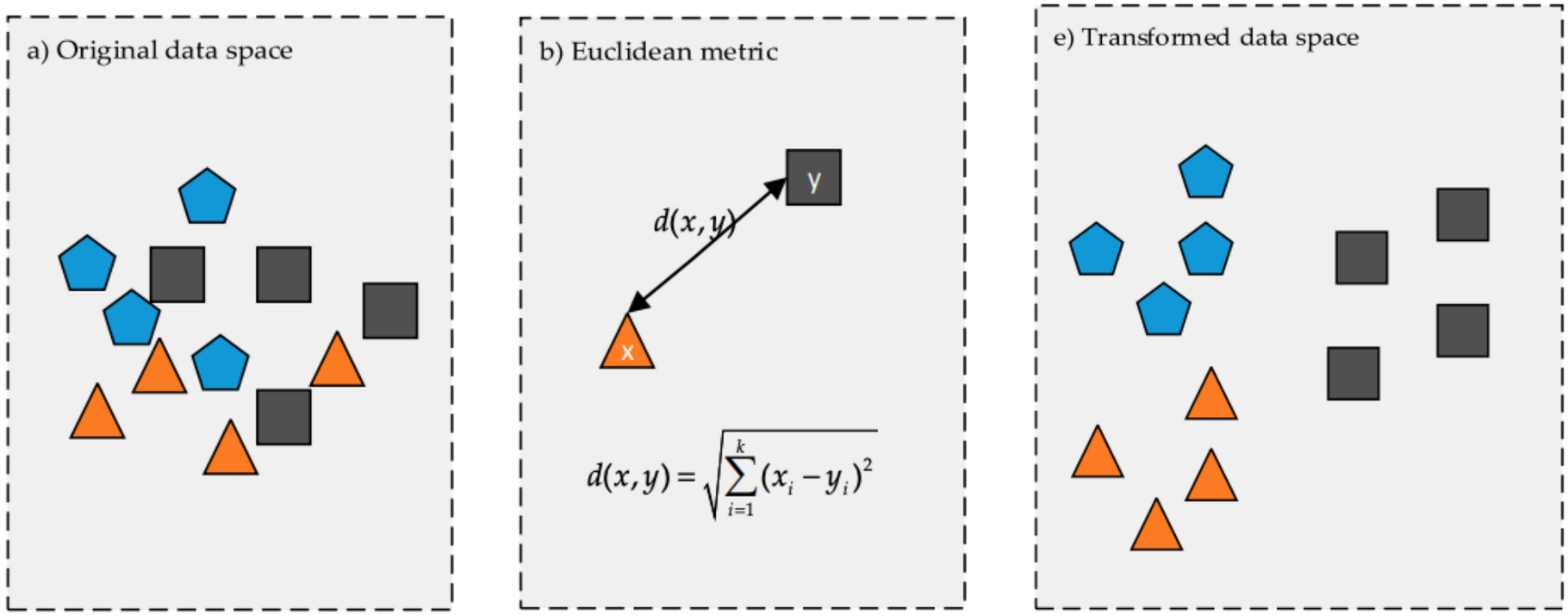


Image Credit: Deep Metric Learning A Survey

Metric Learning

$$d(x,y) = d_A(x,y) = ||x-y||_A = \sqrt{(x-y)^T A(x-y)}.$$

Learn A

Deep Metric Learning

$$d(x,y) = d(f_\theta(x), f_\theta(y))$$

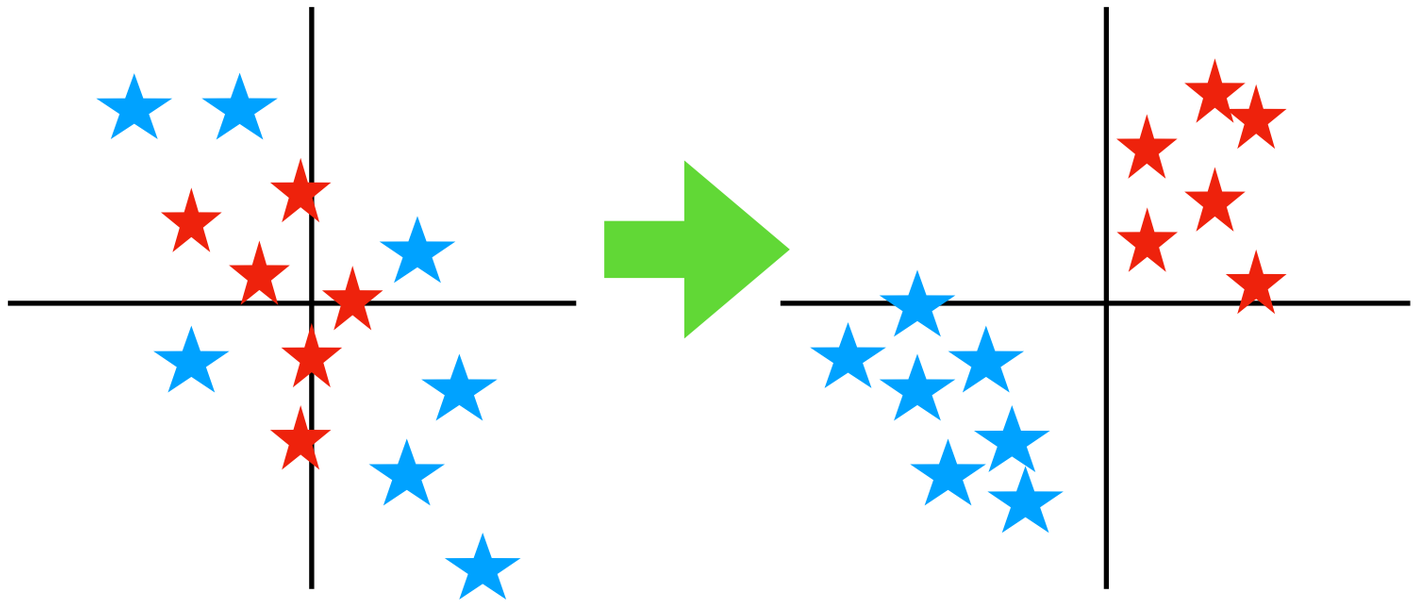
Learn NN f

Distance metric learning, with application to clustering with side-information

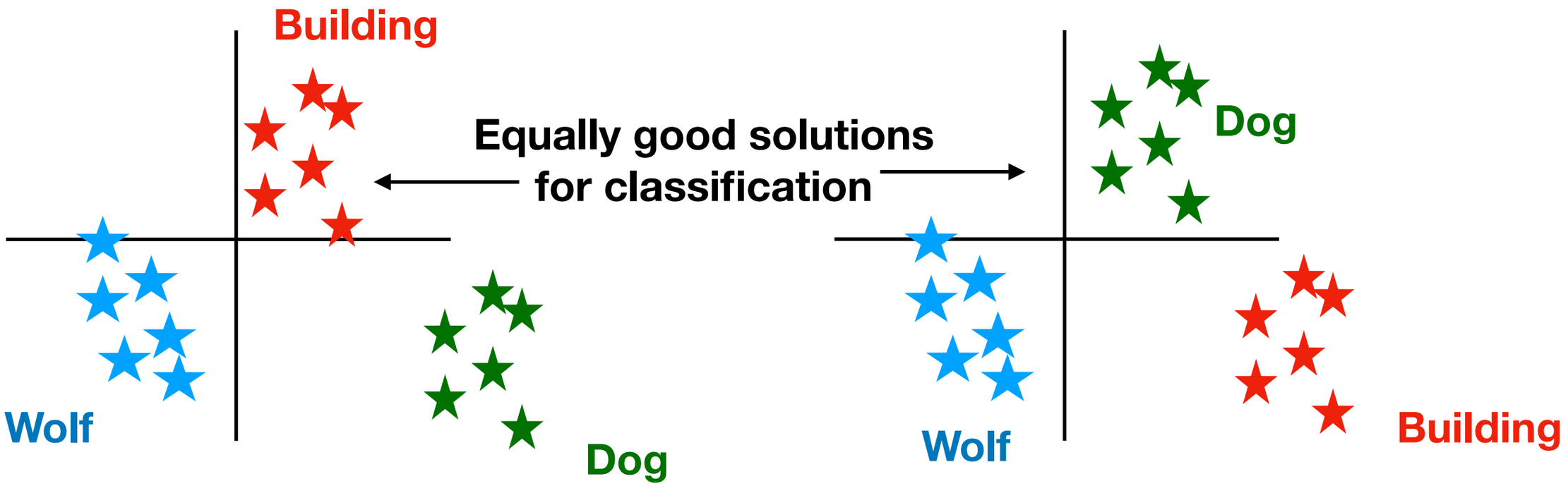
Eric P. Xing, Andrew Y. Ng, Michael I. Jordan and Stuart Russell
University of California, Berkeley

Classification vs Metric Learning

Learned or handcrafted features



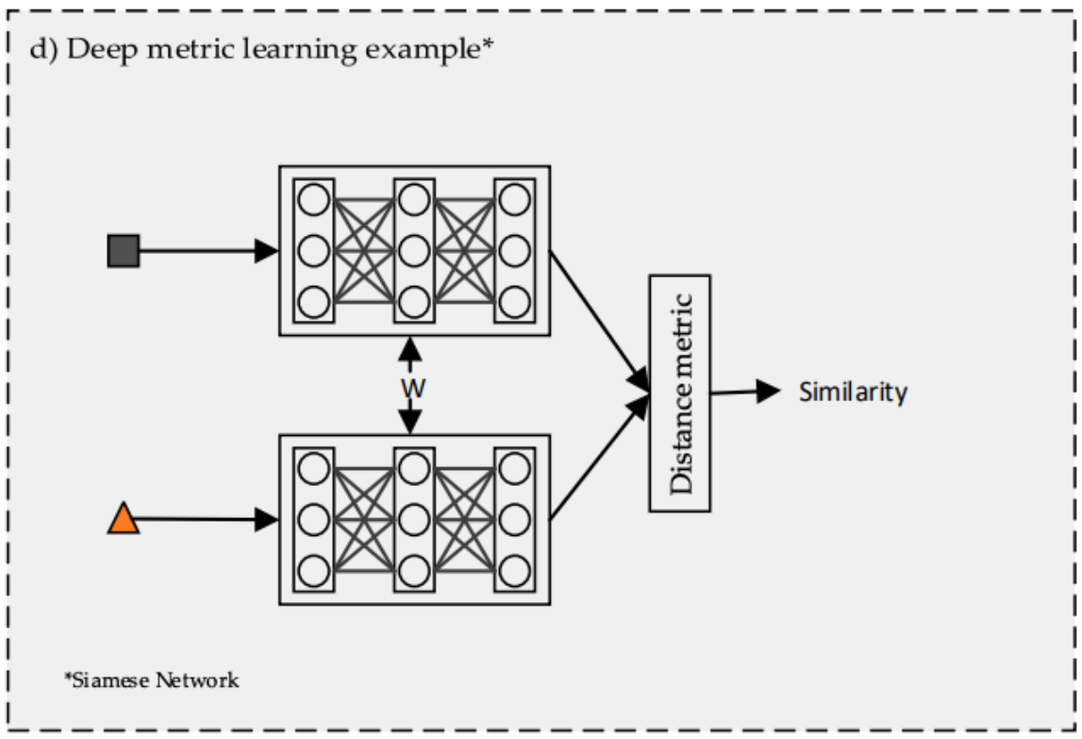
Goal is to get linear separability



Siamese Network

$$d(x, y) = d(G_W(x_1), G_W(x_2))$$

$$e.g. \quad \|G_W(x_1) - G_W(x_2)\|$$

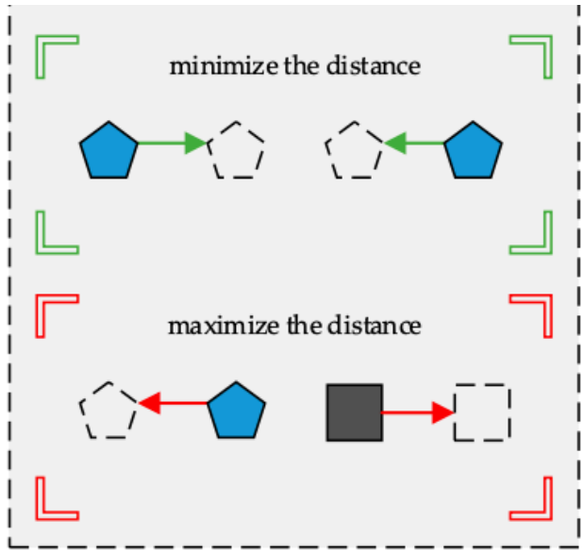


- Consider a classification or “verification” problem, if we just minimize distances between same categories, is it sufficient?

Contrastive (Pairwise) Loss

- Siamese network is typically trained with a pairwise ranking loss often also called contrastive loss

$$z_1 = f_\theta(x_1) \quad z_2 = f_\theta(x_2)$$



$$l(z_1, z_2, y) = y \parallel z_1 - z_2 \parallel + (1 - y)\max(0,\ m - \parallel z_1 - z_2 \parallel)$$

m the margin

$y - >$ **0 or 1 based on whether samples are from same class or not**

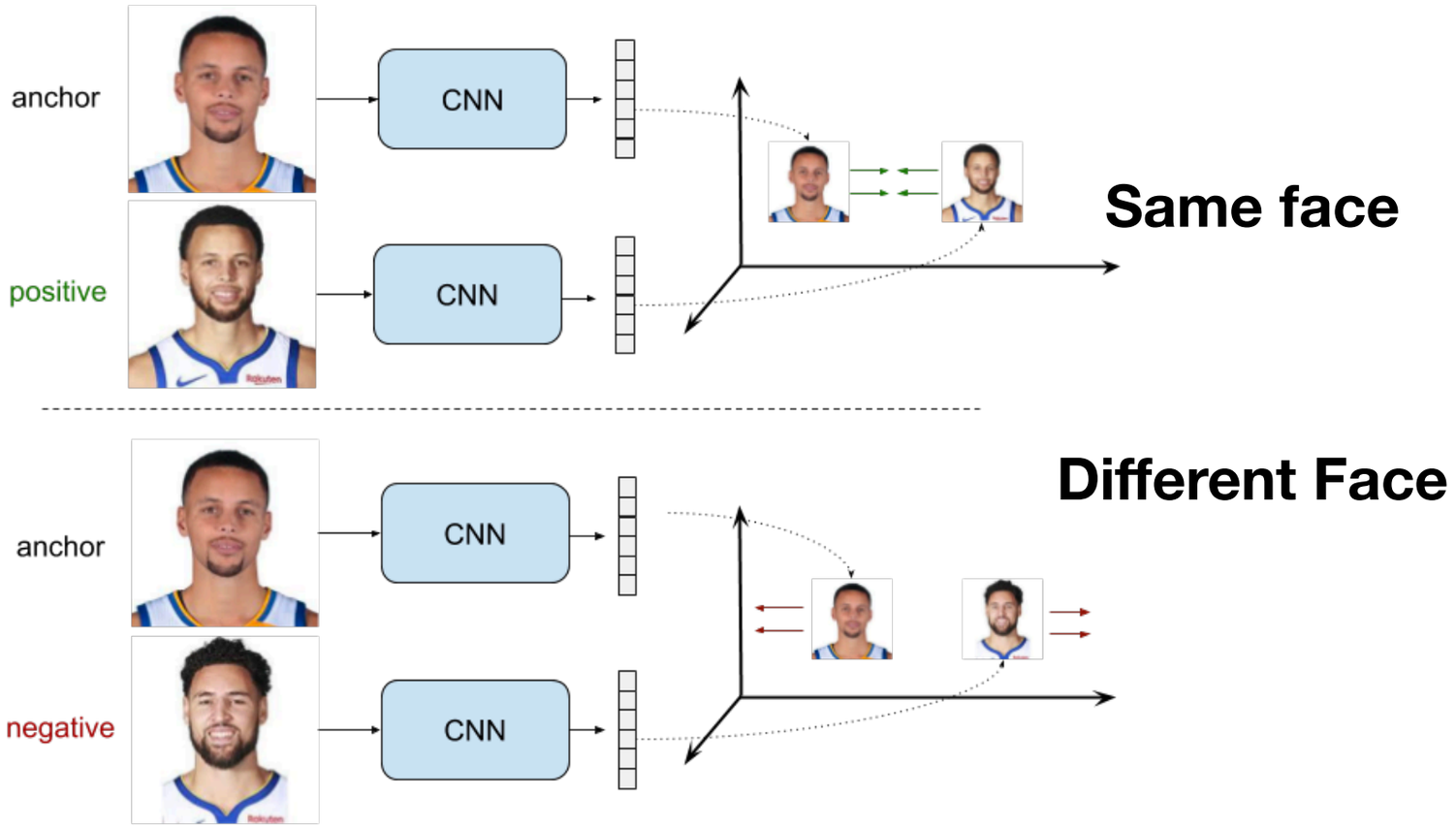
Can we perform updates with just 1 sample as in regular SGD?

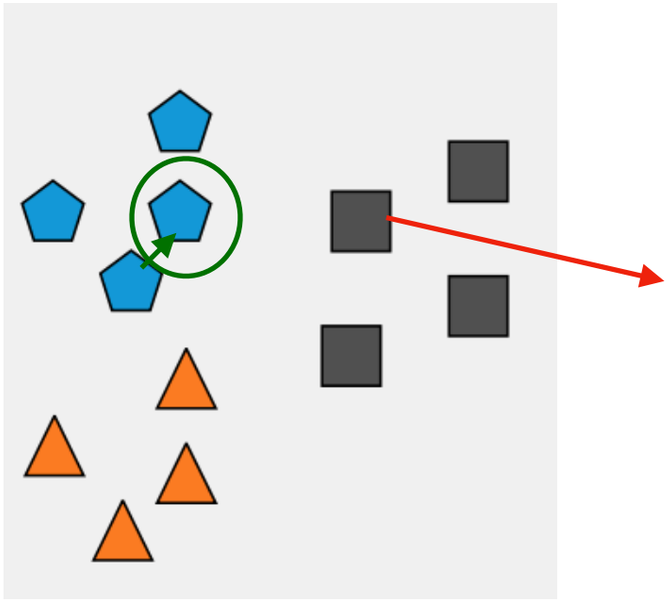
Learning a Similarity Metric Discriminatively, with Application to Face Verification

Sumit Chopra Raia Hadsell Yann LeCun

Issues with Pairwise Loss

$$l(z_1, z_2, y) = y \parallel z_1 - z_2 \parallel + (1 - y)\max(0, m - \parallel z_1 - z_2 \parallel)$$



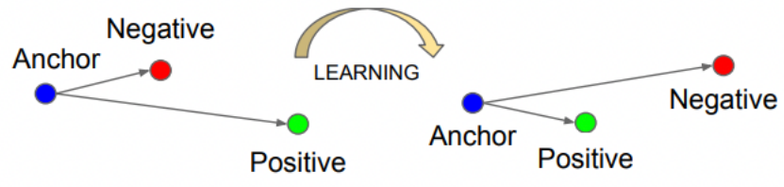


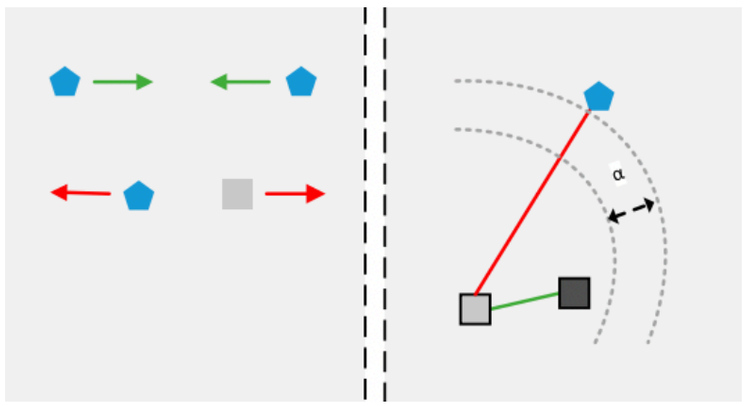
- The simple pairwise loss can help us learn ranks but has issues as it does not enforce relative distances
- Consider the 3 samples above we care about preserving relative distances more than totally satisfying both constraints
- Pairwise loss can be overly aggressive way to do this
  - Same categories can be forced to essentially the same point -> e.g. collapse within class variability
  - Different categories can be forced too far away —> can leave too large distances between categories

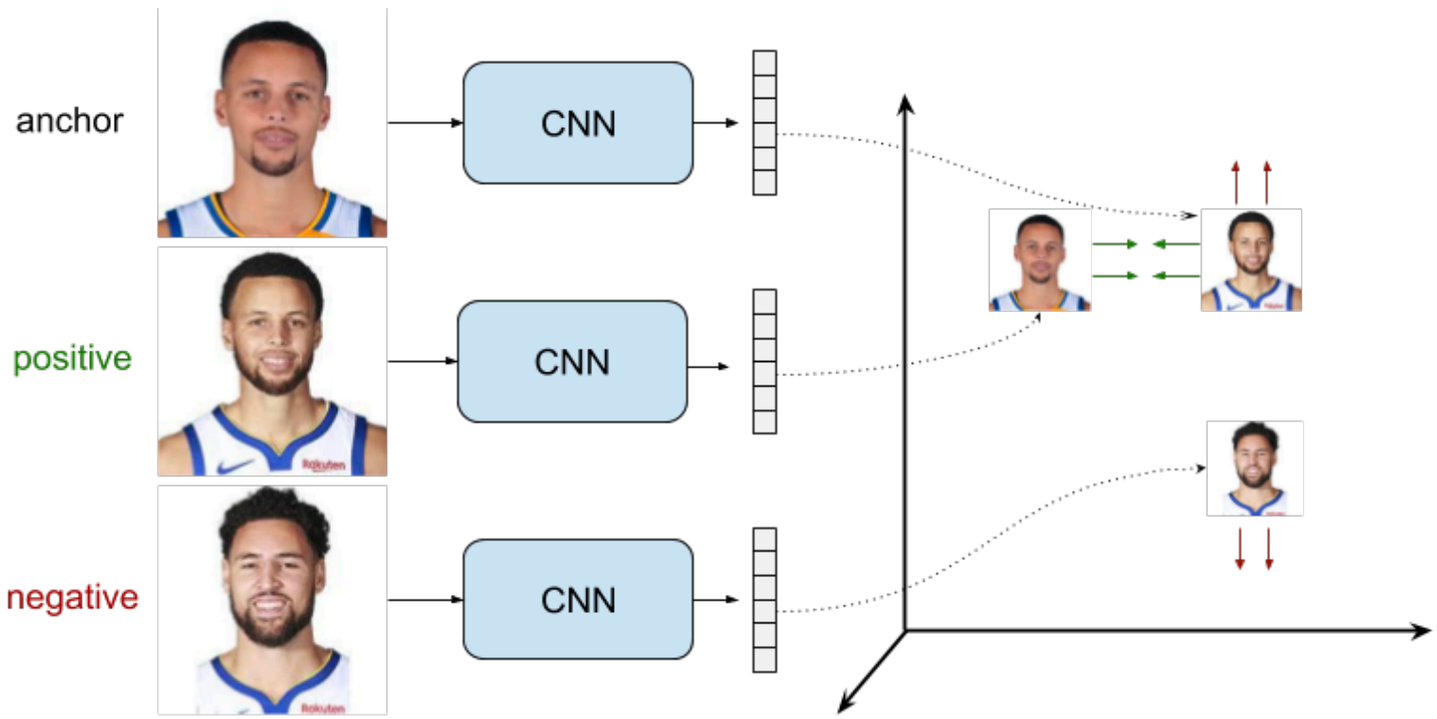
Triplet Loss

$$L(z_a, z_p, z_n) = max(0, m + d(z_a, z_p) - d(z_a, z_n))$$

- Instead of pairwise comparisons we constraint the relative distance to the anchor point:
  - Same category as the anchor should be brought closer
  - Different category as the anchor are pushed away
- Loss now decomposes over triplets versus pairs

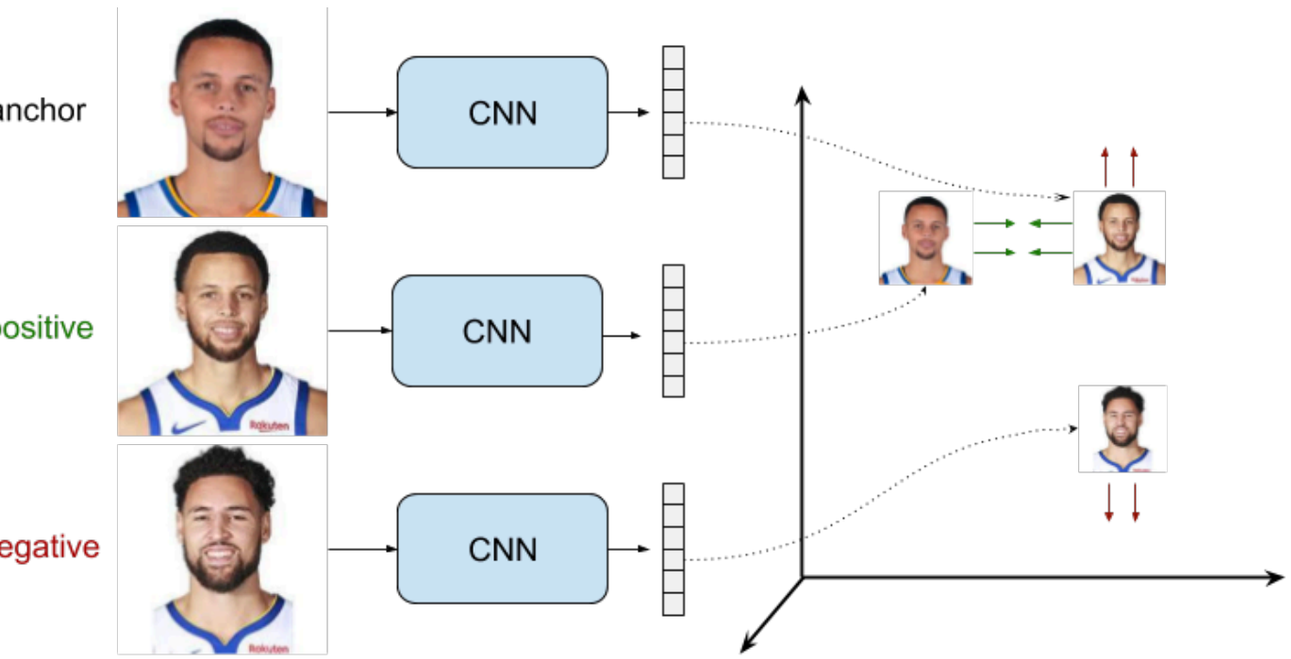






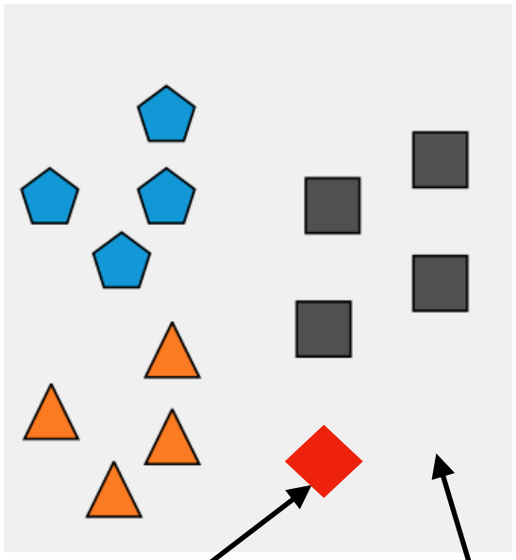
Case Study: Face Verification/Identification

$$L(z_a, z_p, z_n) = max(0, m + d(z_a, z_p) - d(z_a, z_n))$$



Train

- Train a face classifier with triplet loss
- For a new face embed one example of the face
- When “verifying” a user embed the new image and check if distance is above a threshold

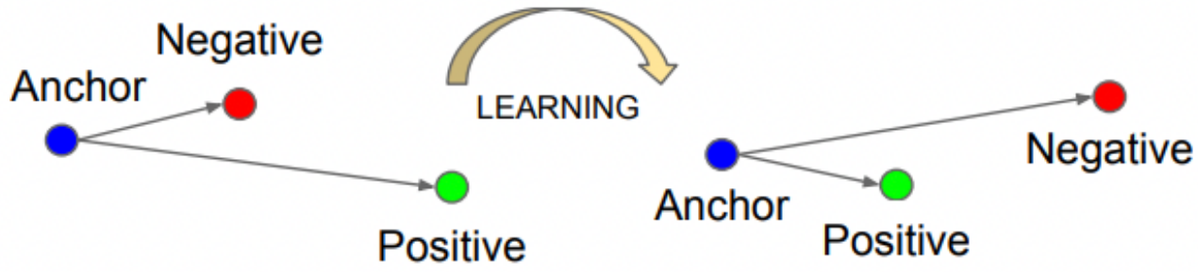


Add New Face

Verify
Or Identify

Triplet Selection

$$L(z_a, z_p, z_n) = max(0, m + d(z_a, z_p) - d(z_a, z_n))$$



- There are a very large number of triplets possible for any dataset

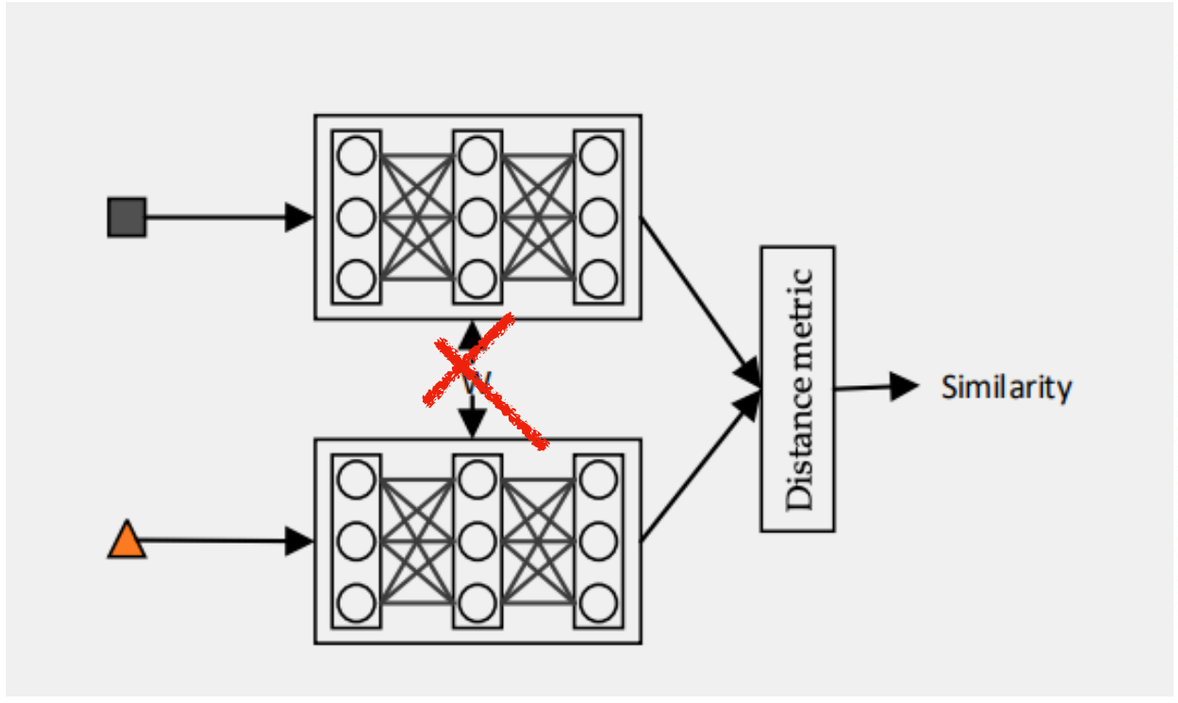
- The model is often already well modelling the distance for certain triplets, further improving these can be ineffective

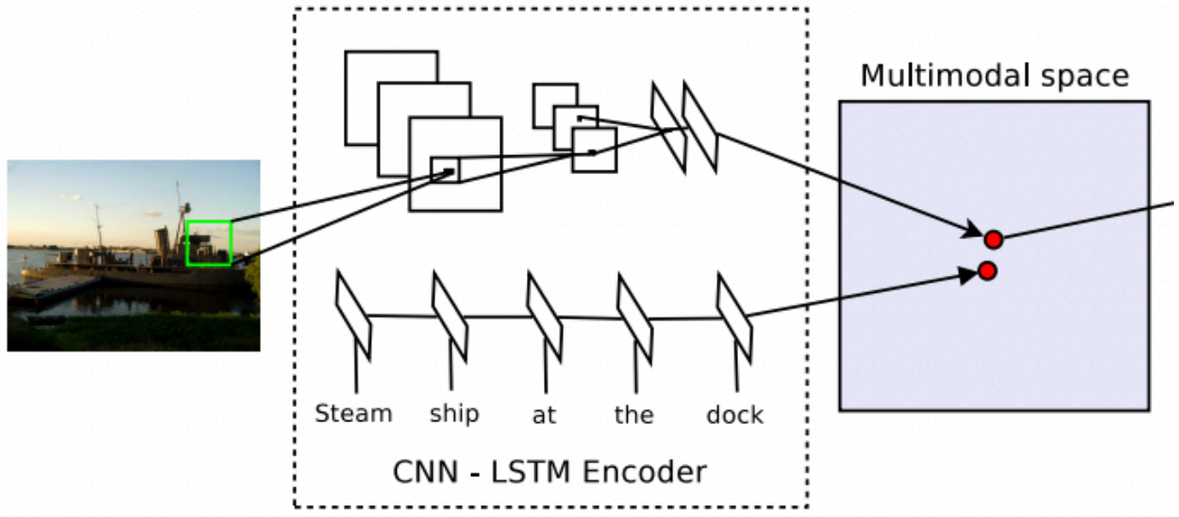
- Hard negative mining tries to find triplets which are currently poorly ranked i.e. $d(r_a, r_n) < d(r_a, r_p)$

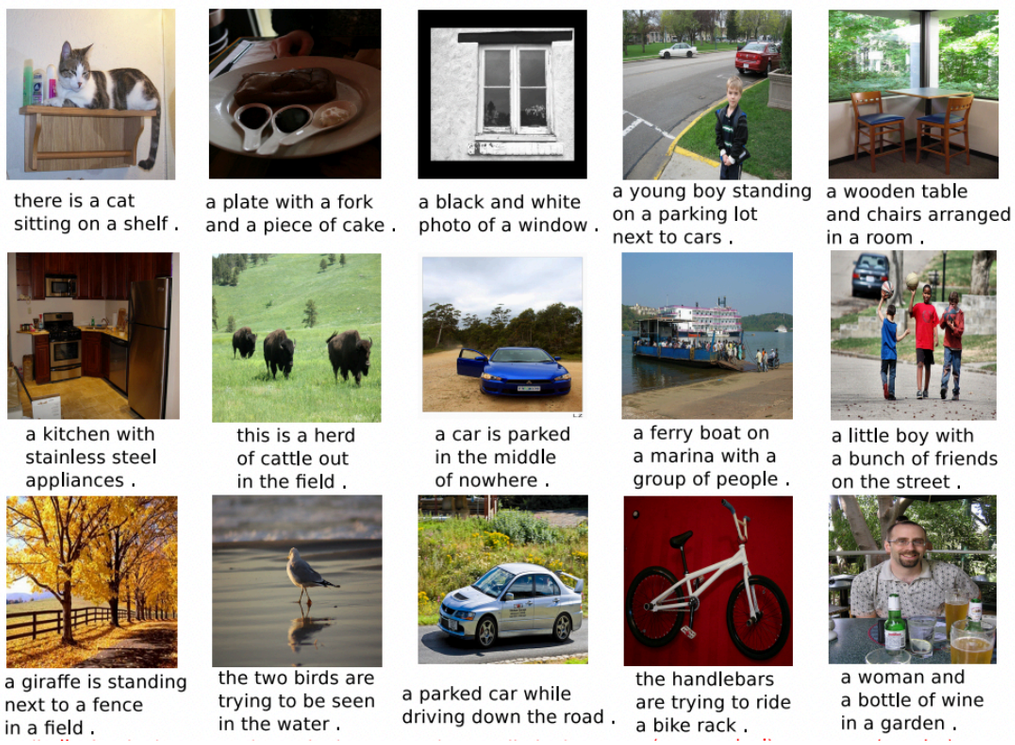
Applications of DML

- Metric Learning methods are well suited for information retrieval tasks
    - e.g. finding relevant documents or similar images
- Metric learning methods are well suited to cases with few samples (e.g. the face verification problem)
    - Once the embedding space is learned adding new classes and performing nearest neighbour can be seen as a non-parametric method
    - Can also be combined with meta-learning (prototypical networks)
- Zero-shot and cross-modal learning

Cross-Modal “Alignment”





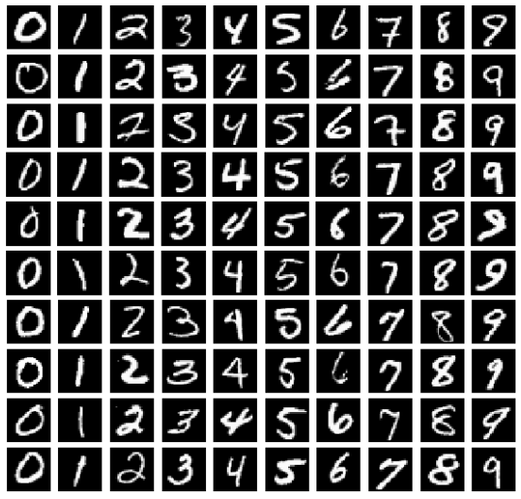


$$d(x, y) = d(f_{\theta_1}(x), g_{\theta_2}(y))$$

Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models

Ryan Kiros, Ruslan Salakhutdinov, Richard S. Zemel
University of Toronto
Canadian Institute for Advanced Research
{rkiros, rsalakhu, zemel}@cs.toronto.edu

Cross-Modal Alignment with Triplet



$$L(z_a, z_p, z_n) = max(0, m + d(z_a, z_p) - d(z_a, z_n))$$

L_1

- Anchor - Domain 1
- Positive - Domain 2 same class as anchor
- Negative - Domain 2 different class as anchor

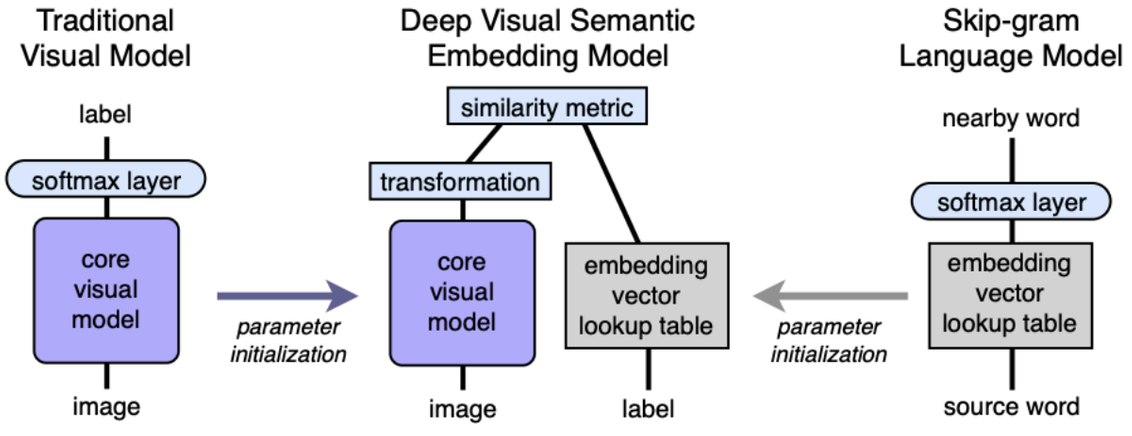
L_2

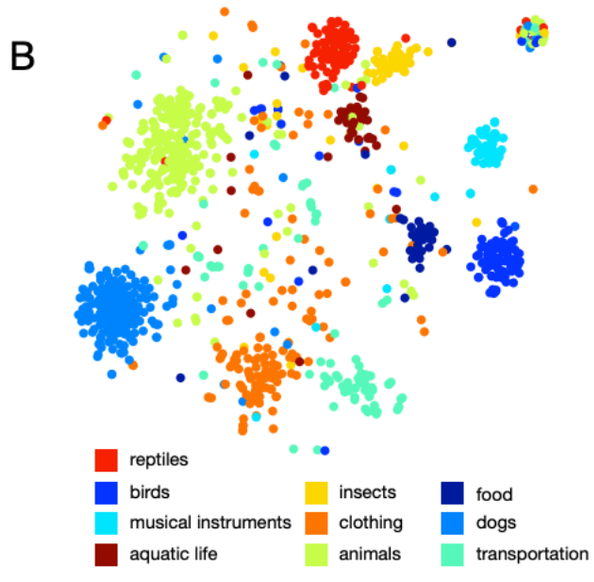
- Anchor - Domain 2
- Positive - Domain 1 same class as anchor
- Negative - Domain 1 different class as anchor

$$L_{total} = L_1 + L_2$$

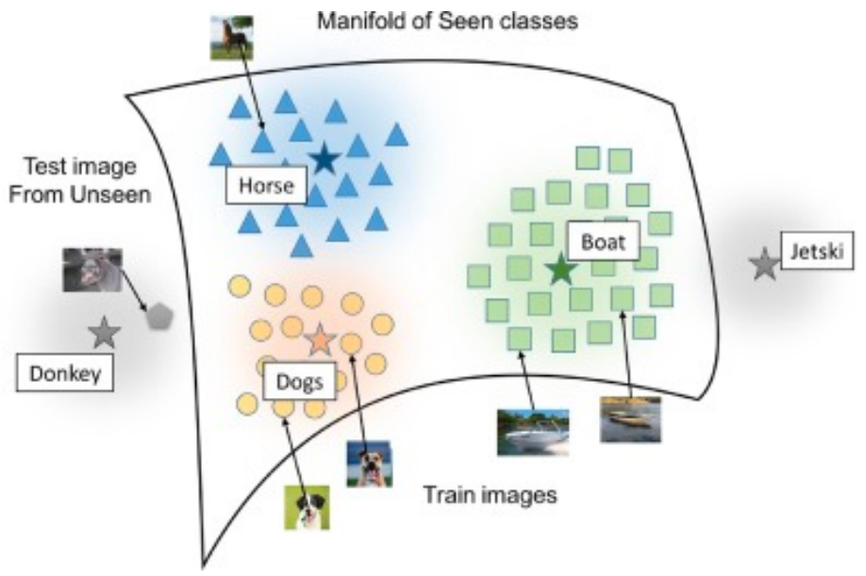
Various ways to construct this in multi-modal setting - can also add classification losses for each modality or within domain triplet loss

Zero-Shot Classification



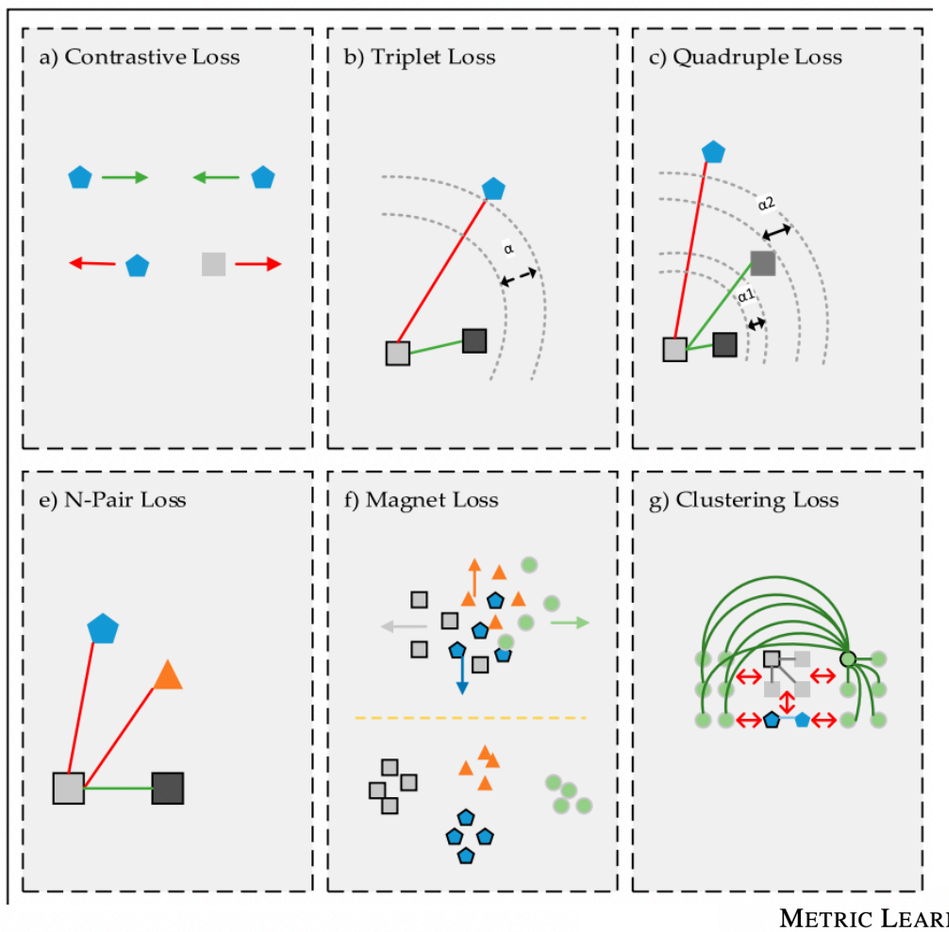


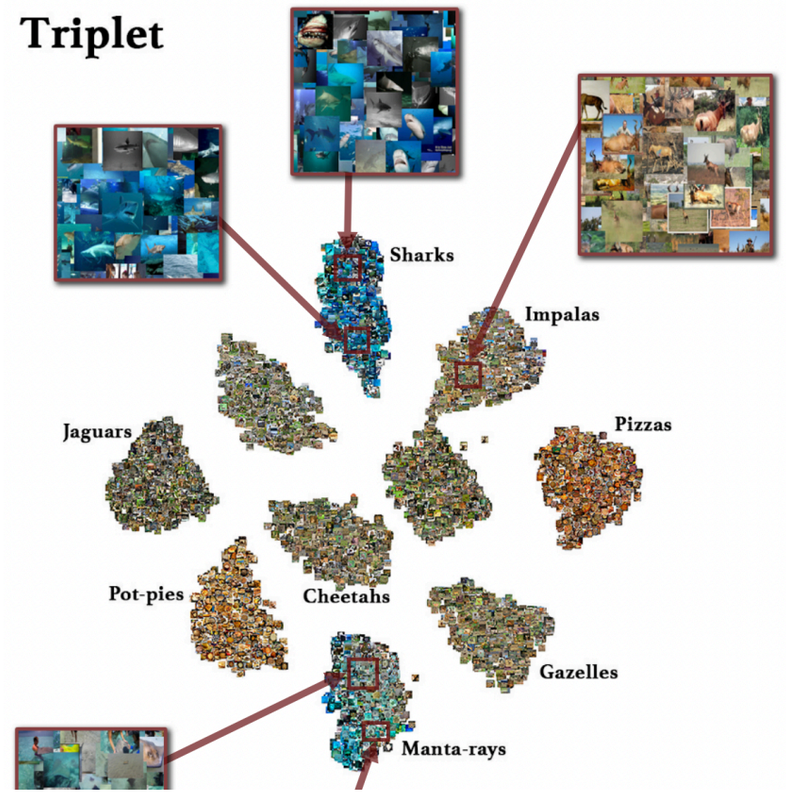
- Multi-modal metric learning allows us to do one form of “zero-shot” classification
- Image to Category example:
  - Train an image model on a set of categories (E.g. 1000 imagenet categories)
  - Trained word embeddings or phrase models (e.g. word2vec) with a broader vocabulary
  - We learn an alignment between the words and images (training time) e.g. with triplet or pairwise loss for example
  - Now at test time we can embed unseen image categories and try to find the nearest word (even words only seen by the language model but not directly by the vision model)
  - More commonly unseen image categories might be combinations of attributes we havent seen in the language model



We don’t need to stop at triplets

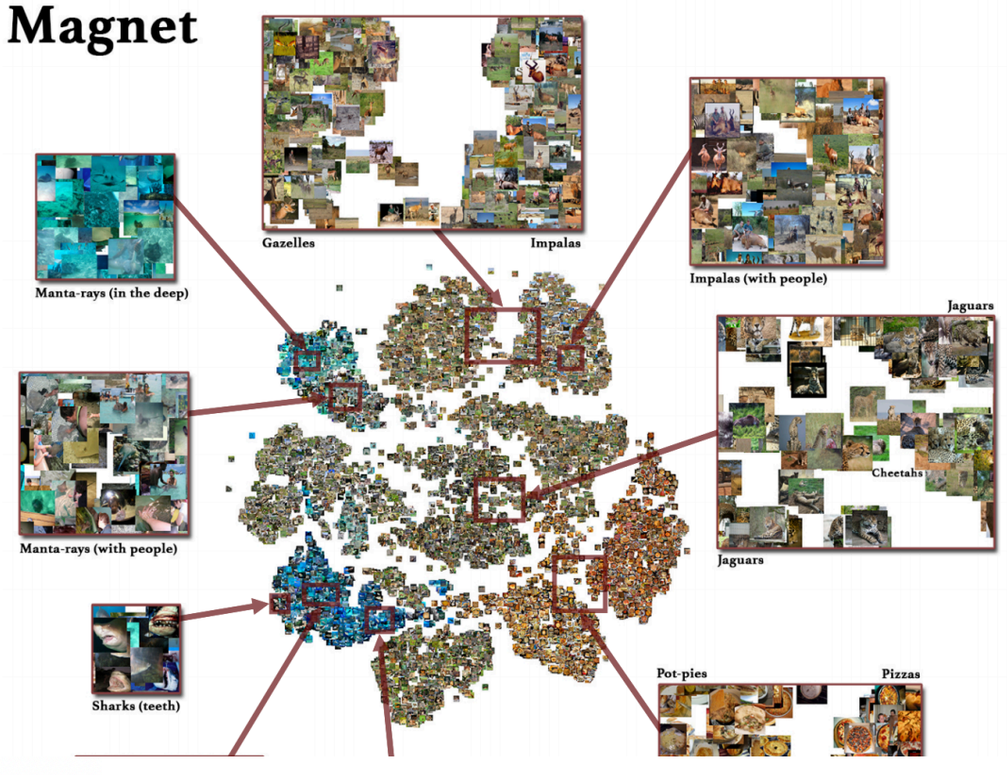
- Various methods that use quadruplets or clustering loss which go across entire batches have been developed
- Can be cumbersome and pairwise or triplet losses still tend to be preferred





$$\mathcal{I}_1^c, \ldots, \mathcal{I}_K^c = \arg\min_{I_1^c,\ldots,I_K^c} \sum_{k=1}^{K} \sum_{\mathbf{r}\in I_k^c} \|\mathbf{r} - \boldsymbol{\mu}_k^c\|_2^2 ,$$

$$\boldsymbol{\mu}_k^c = \frac{1}{|I_k^c|} \sum_{\mathbf{r}\in I_k^c} \mathbf{r} .$$



Metric Learning with Adaptive Density Discrimination

Oren Rippel
MIT, Facebook AI Research
rippel@math.mit.edu

Manohar Paluri
Facebook AI Research
mano@fb.com

Piotr Dollar
Facebook AI Research
pdollar@fb.com

Lubomir Bourdev
UC Berkeley
lubomir.bourdev@gmail.com