Fine-grained Categorization and Dataset Bootstrapping using Deep Metric Learning with Humans in the Loop

Cornell University Department of Computer Science
CORNELL NYCTECH
NEC Laboratories America — Relentless passion for innovation

Yin Cui[1,2] Feng Zhou[3] Yuanqing Lin[3] Serge Belongie[1,2]

[1] Department of Computer Science, Cornell University [2] Cornell Tech [3] NEC Labs America

LAS VEGAS CVPR16

Motivation

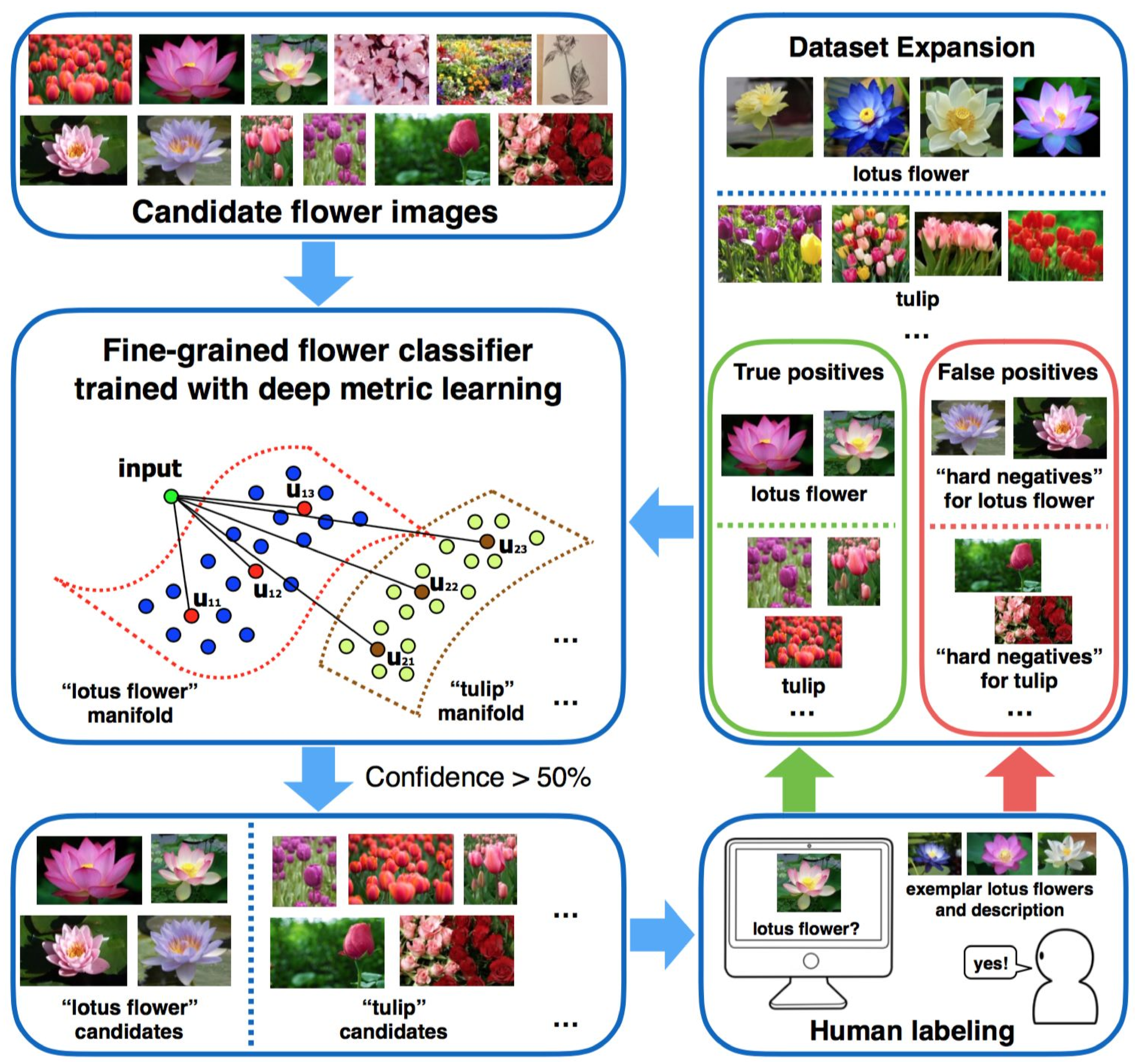
FGVC Challenges

- Lack of training data.
- Large number of categories.
- High intra-class vs. low inter-class variances.

Proposed Solutions

- Bootstrapping training data from the web.
- Learning compact low-dim representations.
- Learning manifolds with multiple anchor points.

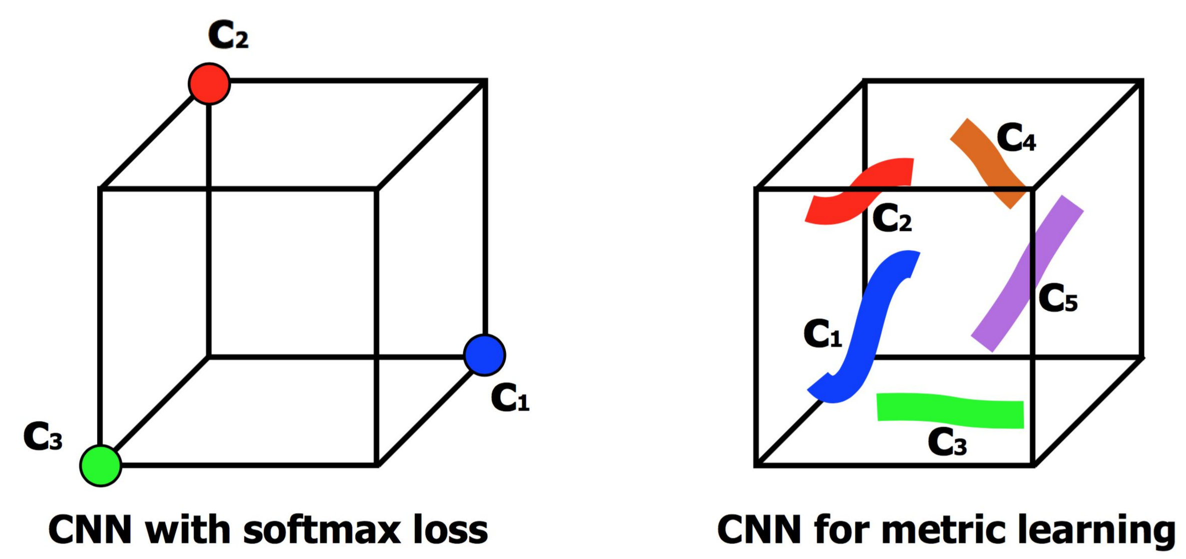
Framework



Contributions

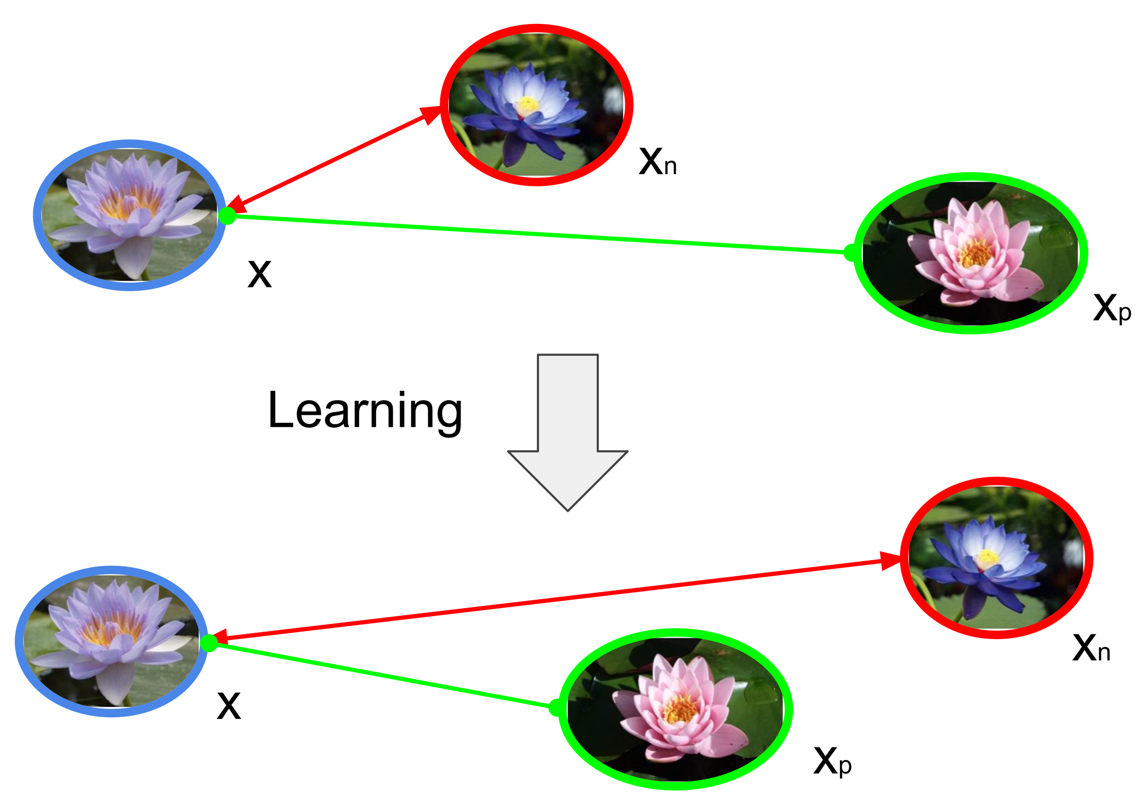
- A **unified framework** for simultaneous *fine-grained categorization and dataset bootstrapping*.
- A novel **metric learning method** that learns manifolds from both *machine-mined* and *human-labeled* hard negatives.
- A **fine-grained flower dataset** with *620 categories* and around *30K images*.

Softmax vs. Metric Learning



- Pre-defined one-hot encoding versus learned manifold.
- Compared with Softmax, metric learning could learn a more *compact* representation in a *much lower dimensional* space.

Triplet-based Metric Learning



- x is more similar to x_p compared with x_n.

$$\mathcal{L}_{triplet}(x, x_p, x_n) =$$
$$\max\left\{0, \|f(x) - f(x_p)\|_2^2 - \|f(x) - f(x_n)\|_2^2 + m\right\}$$

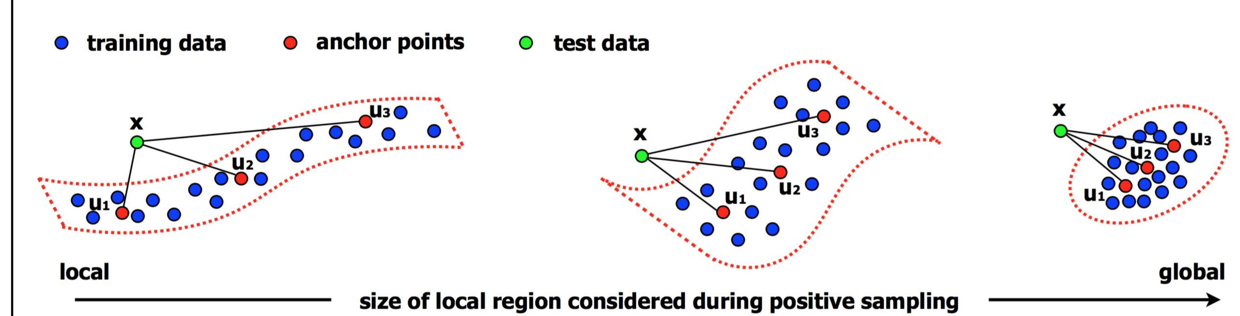
Learning Manifolds

Hard Negatives

- O(n^3) possible triplets, impossible to go through.
  → Need a good sampling strategy.
- Training from hard negatives by:
  a. Only keeping triplets that violate constraint.
  b. Including human-labeled false positives.

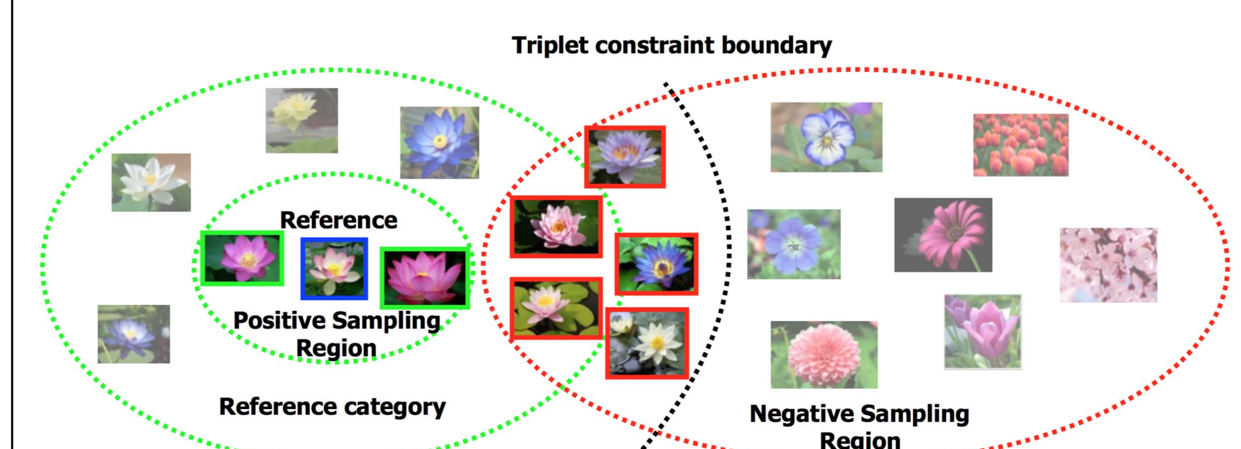
Local Positives

- Sampling local positives could learn a more spread manifold rather than a dense sphere.



Triplet sampling strategy

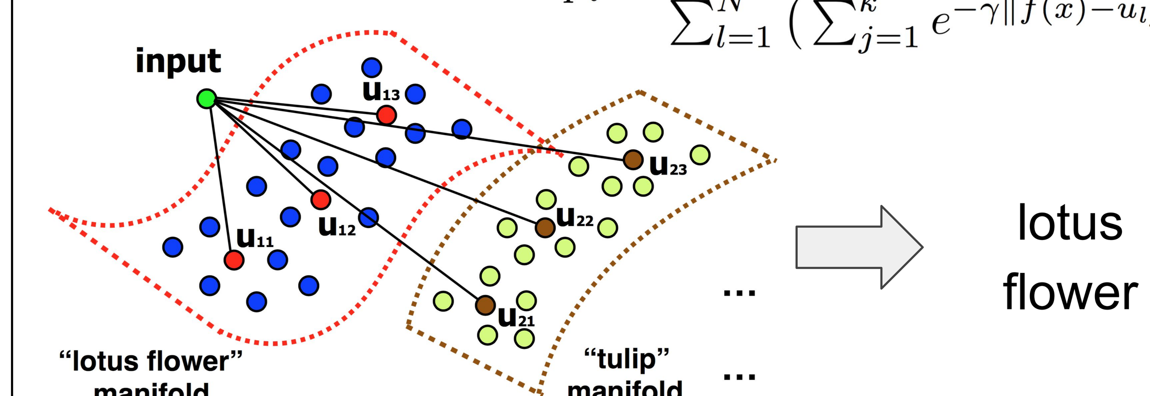
- Hard negatives + local positives.



Classification

- K-means clustering to find anchor points.
- Soft k-nn classifier.

$$p_i = \frac{\sum_{j=1}^{K} e^{-\gamma\|f(x)-u_{ij}\|_2^2}}{\sum_{l=1}^{N}\left(\sum_{j=1}^{k} e^{-\gamma\|f(x)-u_{lj}\|_2^2}\right)}$$

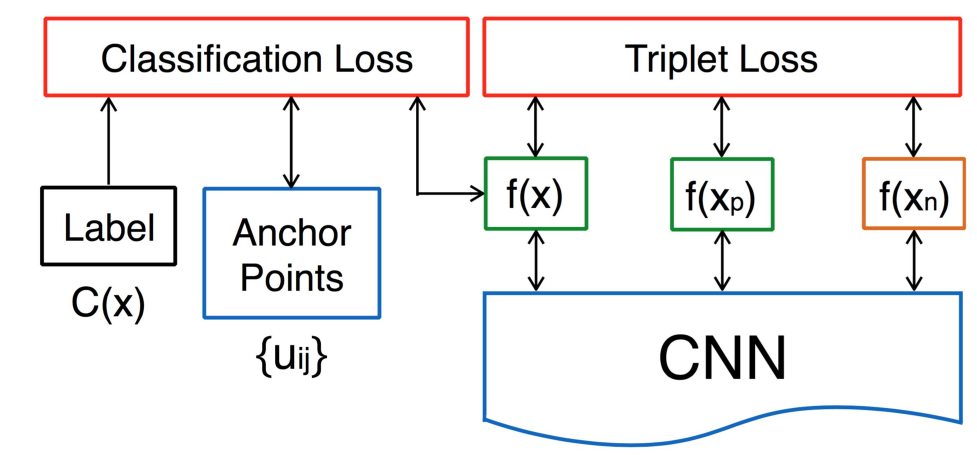


Learning Anchor Points

- Incorporating class labels into metric learning.
- Back-propagate classification loss to update anchor points.

$$\mathcal{L}_{classification}(x, \{u_{ij}\}, C(x)) = -\log(p_{C(x)})$$

$$\mathcal{L} = \omega\mathcal{L}_{triplet} + (1-\omega)\mathcal{L}_{classification}$$



Experiments

Original Flower-620 (15K images)

Method (feature dimension)	Accuracy (%)
Softmax (620)	65.1
Triplet-Naive (64)	48.7
Triplet-HN (64)	64.6
Triplet-M (64)	65.9
Triplet-A (64)	**66.8**

Flower-620 + Instagram images (15K + 15K images)

Method (feature dimension)	Accuracy (%)
Softmax (620)	68.9
Softmax + HNS (621)	70.3
Softmax + HNM (1240)	70.8
Triplet-A (64)	70.2
Triplet-A + HN (64)	**73.7**

naive: random sampling; HN: hard negative mining; M: HN + local positive sampling; A: HN + anchor point learning.

HNS: all human labeled hard negatives as a single category; HNM: human labeled hard negatives for each class as a single category.

- Metric Learning: **+2.7%** over softmax, with a much more compact representation.
- Dataset Bootstrapping: **+6.9%** (**+3.4%** from new data, **3.5%** from human-labeled hard negatives).

Visualization of flower embedding