ACM MM 2010

Image Retagging

Dong Liu, Xian-Sheng Hua, Meng Wang and Hong-Jiang Zhang

Harbin Institute of Technology
Microsoft Research Asia
Microsoft Advanced Technology Center

Social Media and the Associated Tags - Towards Large-Scale Content-Based Multimedia Search

flickr

medici chapel, Firenze, Italy…

Loggia dei lanzi, sword, honeymoon, …

YouTube

Status, building, sky, Italy, …

Cathedral, tower, Italy…

Challenge

Social tags are good, but they are

- Noisy
- Ambiguous
- Incomplete
- No relevance information

Two directions to improve tag quality

- Tag Ranking (Liu, Hua, Zhang. *Tag Ranking*. WWW 09)
- Retagging (Liu, Hua, Wang, Zhang. *Image Retagging*. MM 10)

Tags associated with social images are imprecise, subjective and incomplete.

kitty
top101
boy
young
lovely

Imprecise Tags	
Subjective Tags	
Missing Tags	grass flower cat animal

What we are going to do:

Improve the quality of the tags to better describe content.

But how can we make it?

Automatically.

Two Observations

- Similar images ⟷ similar tags

bear

Visual and Semantic Consistency

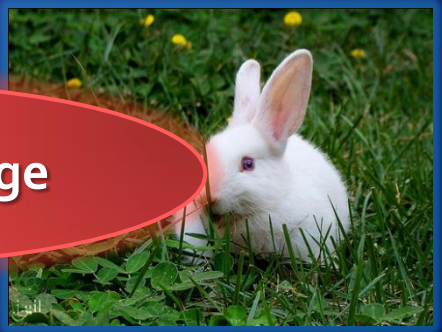
river
bear
animal
bath

- User-provided tags correlate with the image content with high probability

power

Prior Knowledge

father



Nikon
cat
animal
garden
rabit

Basic Assumptions

- Tag Refinement
  - The consistency between visual similarity and semantic similarity should be maximized.
  - The deviation from the initially user-provided tags should be minimized.

Tag Refinement

- Notations

- a social image collection $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$
- All unique tags $\mathcal{T} = \{t_1, t_2, \ldots, t_m\}$
- tag membership $\hat{\mathbf{Y}} \in \{0,1\}^{n \times m}$

if t_j is associated with image x_i, then $\hat{Y}_{ij} = 1$

- refinement results $\mathbf{Y} \quad Y_{ij} \geq 0$
- confidence score vector $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im})^\top$ o image x_i
- visual similarity $W_{ij} = exp(-\frac{\| x_i - x_j \|_2^2}{\sigma^2})$,
- semantic similarity $S_{ij} = \frac{2 * IC(lcs(t_i, t_j))}{IC(t_i) + IC(t_j)}$,

Tag Refinement

- Modeling the basic assumptions

- Visual and semantic consistency

$$\min_{\mathbf{Y}} \sum_{i,j=1}^{n} (W_{ij} - \sum_{k,l=1}^{m} Y_{ik} S_{kl} Y_{jl})^2,$$
$$s.t. \quad Y_{jl} \geq 0, \quad j = 1, 2, \ldots, n, \quad l = 1, 2, \ldots, m.$$

- User-provided tags are relevant with high probability

$$\sum_{j=1}^{n} \sum_{l=1}^{m} (Y_{jl} - \alpha_j \widehat{Y}_{jl})^2 \exp(\widehat{Y}_{jl}).$$

- Overall formulation

$$\min_{\mathbf{Y}, \boldsymbol{\alpha}} \quad \mathcal{L} = \sum_{i,j=1}^{n} (W_{ij} - \sum_{k,l=1}^{m} Y_{ik} S_{kl} Y_{jl})^2$$
$$+ C \sum_{j=1}^{n} \sum_{l=1}^{m} (Y_{jl} - \alpha_j \widehat{Y}_{jl})^2 \exp(\widehat{Y}_{jl}),$$
$$s.t. \quad Y_{jl}, \alpha_j \geq 0, \quad j = 1, 2, \ldots, n, \quad l = 1, 2, \ldots, m.$$

Tag Refinement

- Optimizing with iterative updating

  - Bound the objective function

$$\mathcal{L} \le \mathcal{L}' = \sum_{i,j=1}^{n} \Big(W_{ij}^2 + \sum_{l=1}^{m} [\tilde{\boldsymbol{Y}}\boldsymbol{S}\tilde{\boldsymbol{Y}}^\top]_{ij} [\tilde{\boldsymbol{Y}}\boldsymbol{S}]_{il} \frac{Y_{jl}^4}{\tilde{Y}_{jl}^3} - 4 \sum_{l=1}^{m} W_{ij} [\tilde{\boldsymbol{Y}}\boldsymbol{S}]_{il} \tilde{Y}_{jl} \log Y_{jl} - 2 W_{ij} [\tilde{\boldsymbol{Y}}\boldsymbol{S}\tilde{\boldsymbol{Y}}^\top]_{ij} + 4 \sum_{k=1}^{m} W_{ij} [\boldsymbol{S}\tilde{\boldsymbol{Y}}^\top]_{kj} \log \tilde{Y}_{ik} \Big) + C \sum_{j=1}^{n} \sum_{l=1}^{m} \Big(Y_{jl}^2 - 2\alpha_j \hat{Y}_{jl} \tilde{Y}_{jl} (\log \frac{Y_{jl}}{\tilde{Y}_{jl}} + 1) + \alpha_j^2 \hat{Y}_{jl}^2 \Big) \exp(\hat{Y}_{jl}),$$

  - Derive the solution

$$\begin{cases} Y_{jl} = \left[\frac{-C \exp(\hat{Y}_{jl}) \tilde{Y}_{jl}^3 + \sqrt{M}}{4 [\tilde{\boldsymbol{Y}}\boldsymbol{S}\tilde{\boldsymbol{Y}}^\top \tilde{\boldsymbol{Y}}\boldsymbol{S}]_{jl}} \right]^{\frac{1}{2}}, \\ \alpha_j = \frac{\sum_{l=1}^{m} \tilde{Y}_{jl} (\log Y_{jl} - \log \tilde{Y}_{jl} + 1)}{\sum_{l=1}^{m} \tilde{Y}_{jl}}, \end{cases}$$

  - Iterative updating until convergence

    - Fix $\boldsymbol{\alpha}$, update $\mathbf{Y}$ using the upper equation
    - Fix $\mathbf{Y}$, update $\boldsymbol{\alpha}$ using the lower equation

Is It Reliable ?

Visual Property of Tags

Content -Related Tag

Describe the **REAL** visual content of the images.
Informative for ALL general users.

Content -Unrelated Tag

Describe the **CONTEXTUAL** information about the images .
Only informative to the image owners.

Our Basic Assumption

RECALL

- Similar images have similar tags.

Only applicable for "content-related" tags.

- Involving the content-unrelated tags will
  - Introduce lots of noises.
  - Degrade the algorithmic performance.

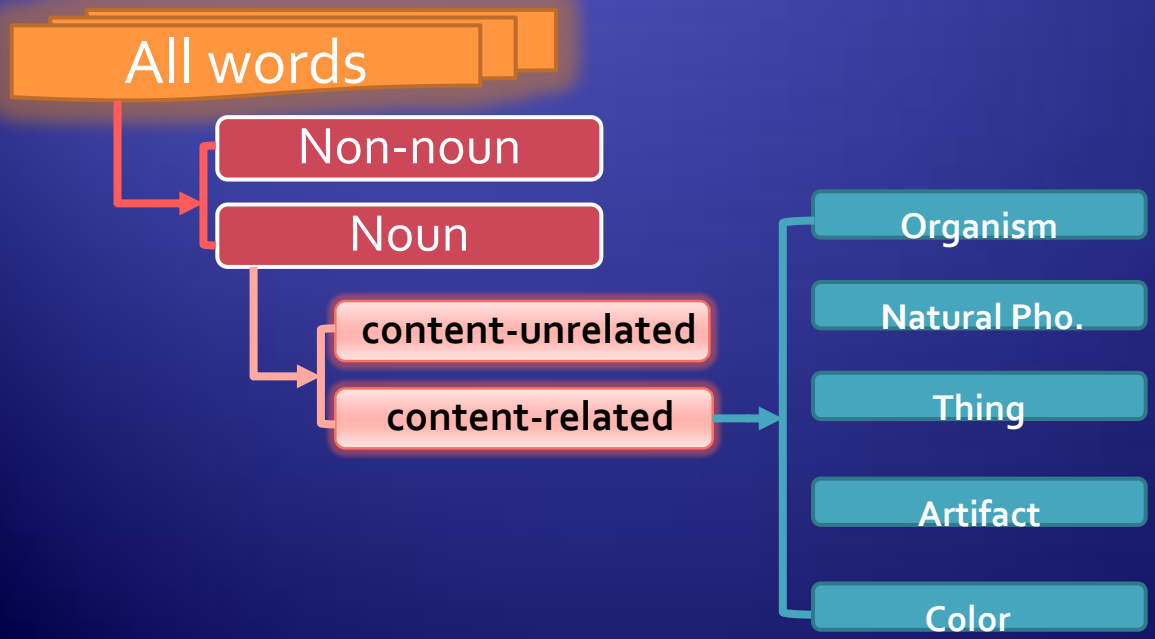
These tags should be removed from the automatic learning procedure.

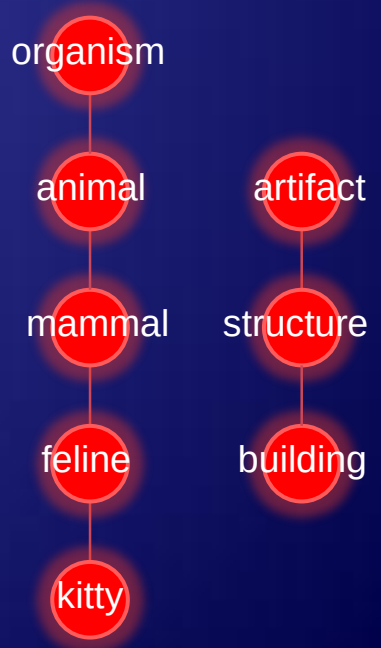
Tag Filtering

- Filter out all content-unrelated tags.

Construct a content-related tag dictionary by using the lexical and domain knowledge

Traverse along the path until one pre-defined category is matched





Is It Enough?

Synonym and hypernym

kitty

kitty

kitten
cat
pussy

synonym

kitty

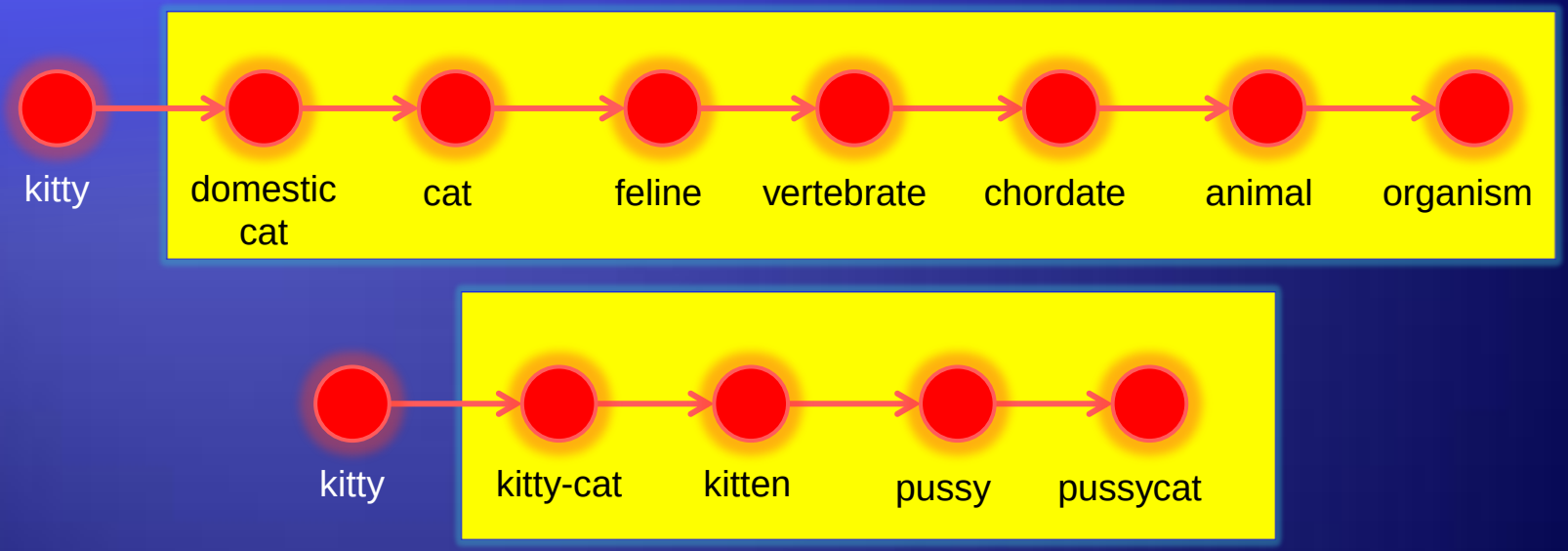
feline
animal

hypernym

The missing of such tags will degrade the performance of tag-based applications.

Further Enrichment

- Make use of Wordnet lexicon

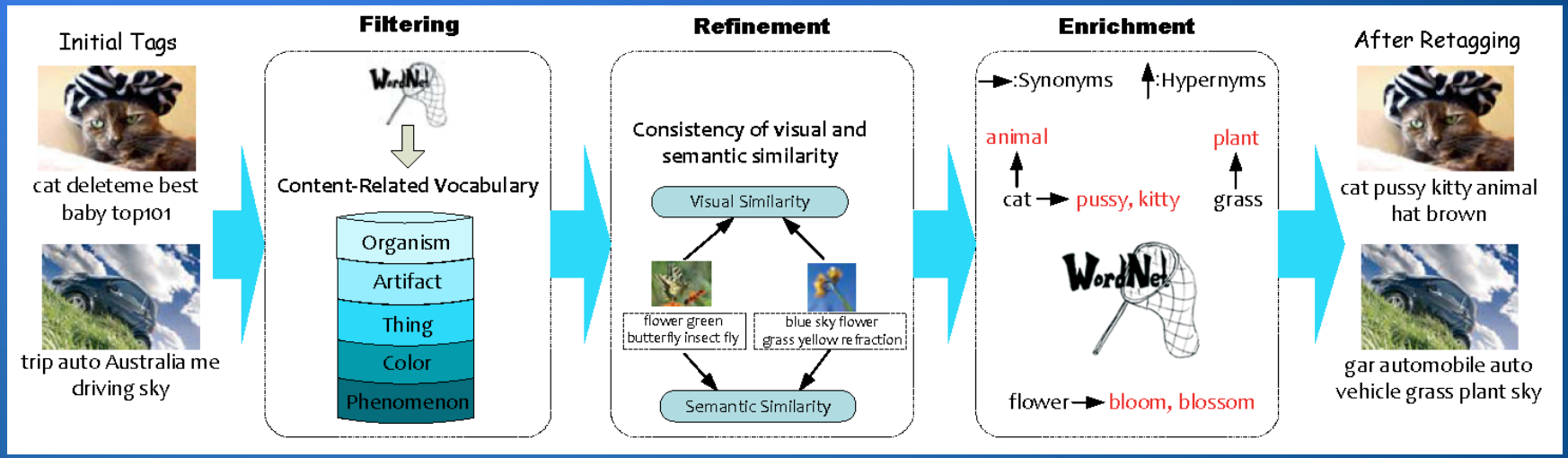


Use each tag to perform tag-based image search on Flickr.

The tags with more than 10,000 returned images are retained.

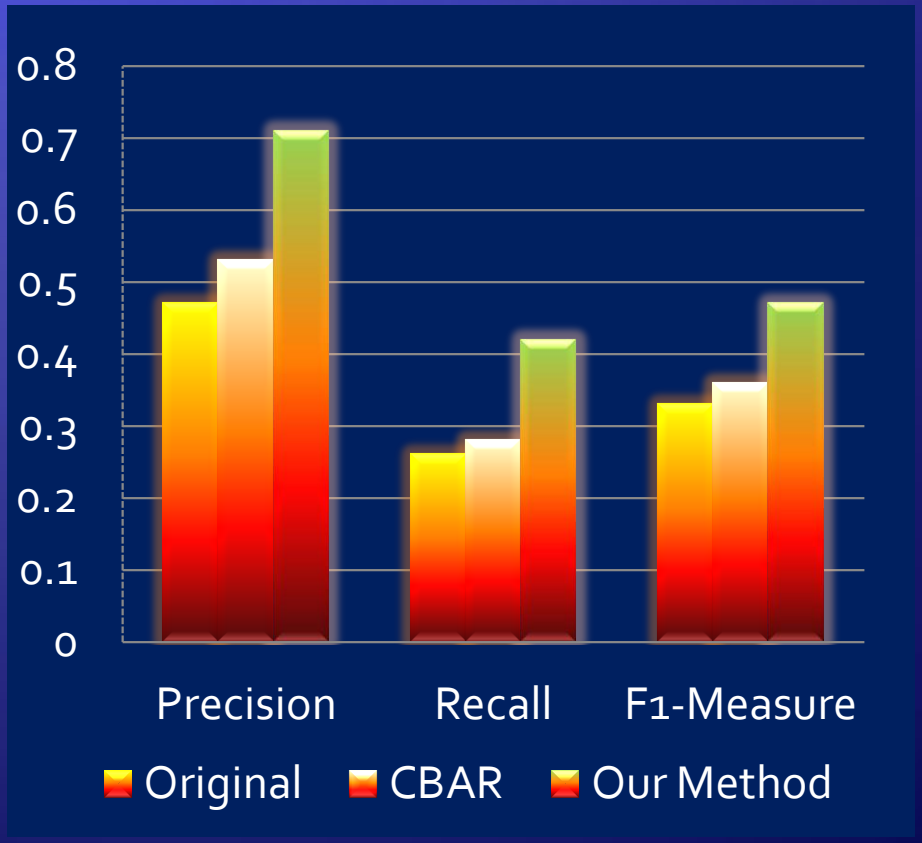
In Summary: Image Retagging

- Three-step strategy



Performance of Tag Refinement

- In term of average precision,recall and F1-Measure
  - 50,000 Flickr images with 4,556 content-related tags.
  - 2,500 test images.



Initial Tags: sunset orange usa boat northcarolina fv10 lakecrabtree morrisville massimostrazzeri nikonstunninggallery

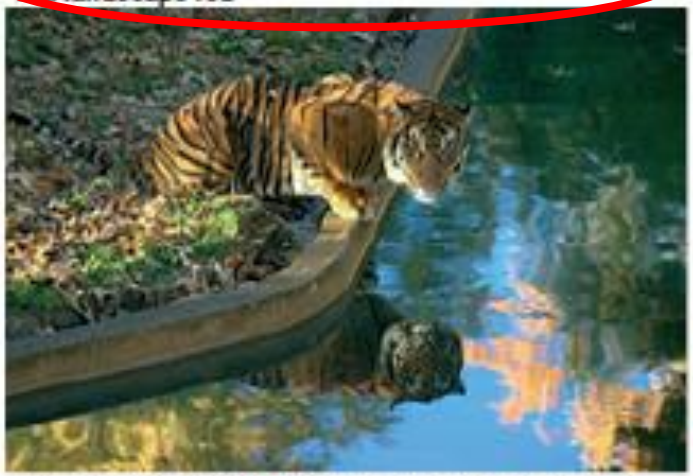
CBAR: water sky white blue nature sunset sun yellow light red

Our Method: sunset sky water nature boat sun flower sea landscape red

Initial Tags: canada reflection bird water birds 1025fav reflections geese bc britishcolumbia

CBAR: blue white water sky nature green yellow nikon tree light

Our Method: water bird sky sunset yellow nature blue cat flower beach



Initial Tags: 2005 cats topf25 animal animals topv111 cat zoo washingtondc smithsonian

CBAR: green nature cat flower landscape sea mountain sky water animal

Our Method: cat sky animal water sunset nature blue zoo tree tiger

Initial Tags: nepal mountain mountains trek wow dawn village annapurna nov99 gandruk

CBAR: sky water blue white nature sunset sun light green yellow

Our Method: mountain sky water sunset nature blue flower dawn tree landscape

Performance of Tag Enrichment

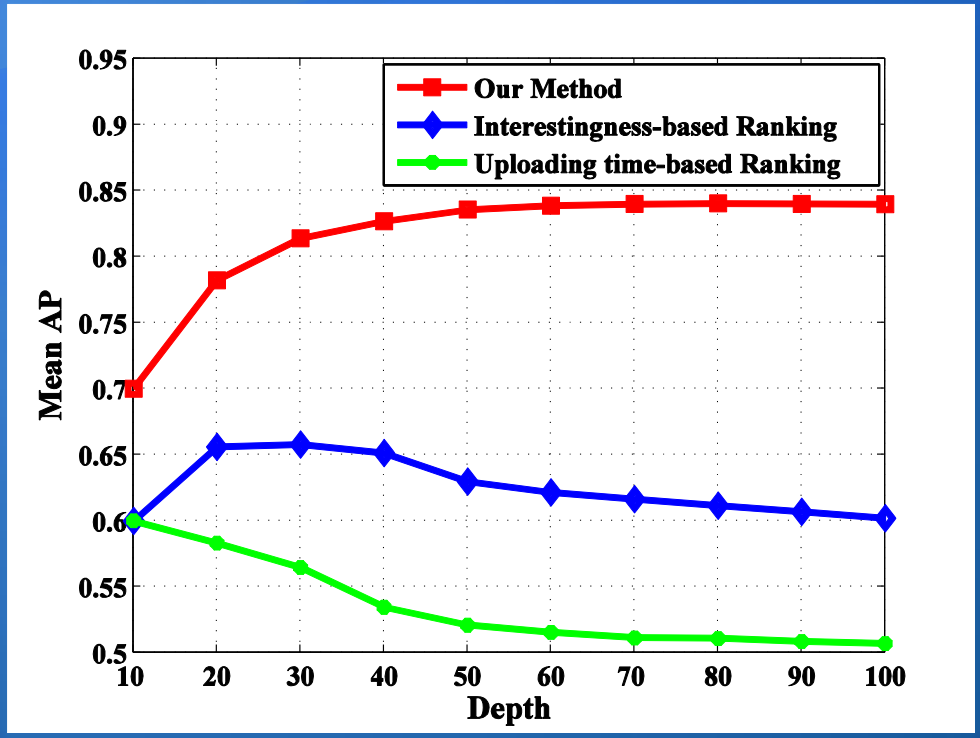
Method	Precision	Recall	F1-measure	Relevant tag num
Before Enrichment	0.71	0.34	0.46	3.09 (4,80 in all)
After Enrichment	0.90	0.66	0.76	9.34 (10.38 in all)

Tagging quality is further improved rafter the tag enrichment procedure.

Application 1: Tag-Based Image Search

- Use the learnt confidence scores as relevance measure
- Ranking results for query "cat"

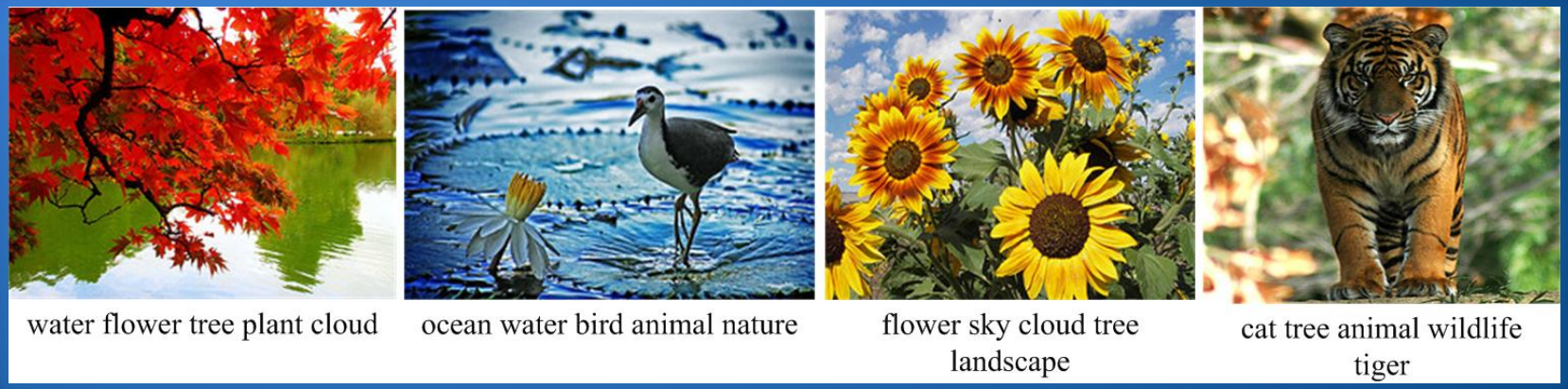
Performance of Tag-Based Search



Our confidence score based ranking strategy outperforms the other image ranking strategies on Flickr

Application 2: Auto Tagging

- Use top tags of the images after retagging to predict the tags of the unlabeled images



Performance of Auto Tagging

Methods	Using images with original tags	Using retagged images
Precision in average	0.221	0.340

Using top tags after image retagging can obtain better results than using the original images directly

Conclusion

- User-provided tags are imprecise and incomplete, which limits the performance of tag-based applications
- We propose an image retagging strategy to solve this problem:
  - Tag filtering to remove the content-unrelated tags
  - Tag refinement to automatically refine the tags
  - Tag enrichment to expand the tags with synonyms and hypernyms.
- Image retagging benefits a series of tag-based applications

Future work

- Extend it to online videos
- Using more fruitful information cues such as image regions and surrounding texts

Thank You