第14回コンピュータビジョン勉強会2011年7月31日

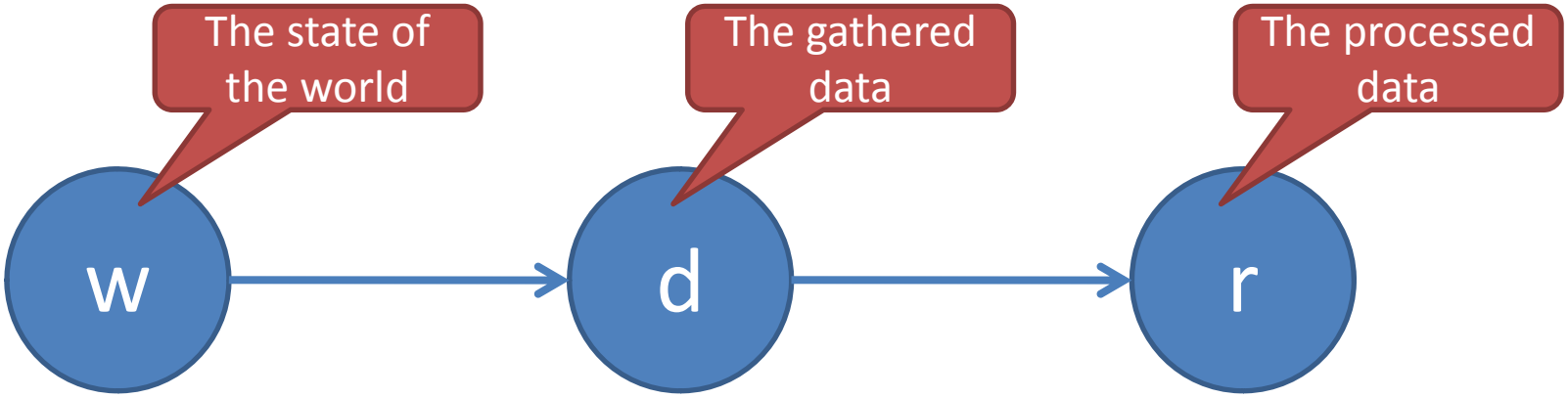
CVPR2011における
一般物体・シーン認識のトレンド

東京大学/JSTさきがけ
原田達也

- 氏名
  - Lena Soderberg
- 生年月日
  - 1951年3月31日
- 現年齢
  - 60歳
- 出身地
  - スウェーデン

The famous single-image-dataset
One of the first “real” image

A. Torralba, A. A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.



Markov chain

The d

$$P(w,d,r) = P(w)P(d\,|\,w)P(r\,|\,d)$$

The average information

$$I(W;D) \geq I(W;R)$$

The data processing theorem states that data processing can only destroy information.

David J.C. MacKay. Information Theory, Inference, and Learning Algorithms. Cambridge University Press 2003.

画像認識のプロセス

訓練時

識別時

- 処理を重ねる毎にデータの持つ情報は減少する.
  - データ，特徴抽出，モデルの順に高い質が求められる.
- 従来の画像認識研究の多くはモデル化に重点が置かれていた
  - 小さな実験環境，スモールワールド
- 複雑なモデルは大規模データの前では役に立たない
  - スケーラビリティの重要性
- 高い質のデータ，特徴抽出が適切に行われていればシンプルなモデルで十分な性能が出せる

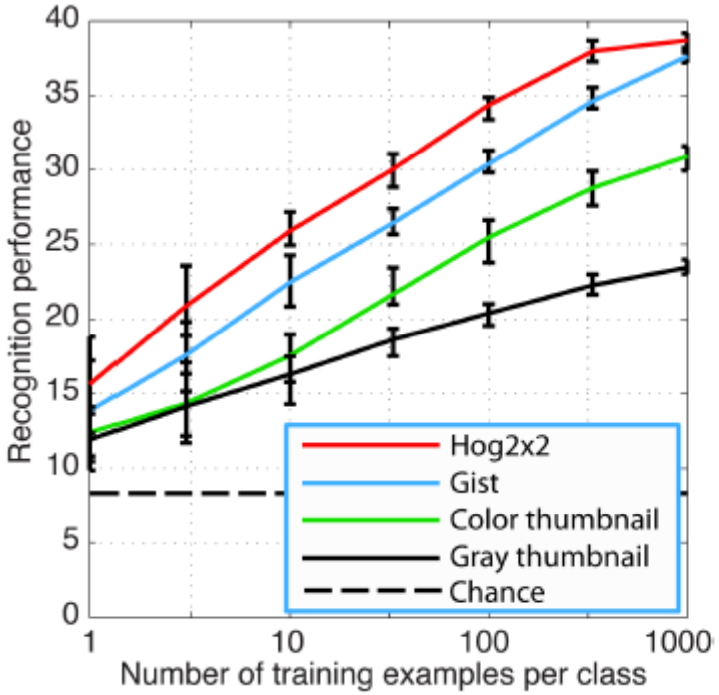
Name That Dataset!

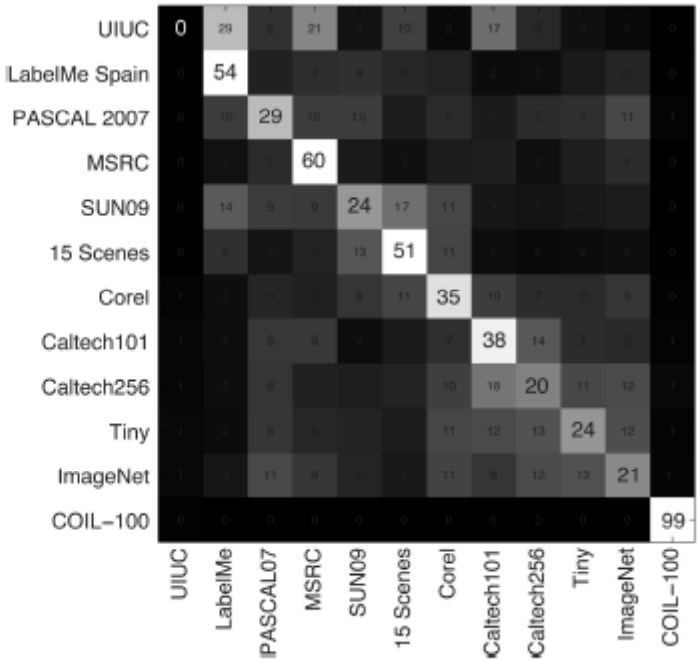
- Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

理論的には困難な課題のはず.
しかし，物体・シーン認識の研究者には比較的易しい課題

Name That Dataset識別機

Antonio Torralba, Alexei A. Efros.
Unbiased Look at Dataset Bias.
CVPR, 2011.





- 特徴量
  - Gray tiny image, color tiny image, gist, BoF (HOG)
- 識別機
  - 線形SVM
- プロトコル
  - 各データセットから1000枚の訓練画像，300枚のテスト画像

なるべく偏り（bias）がないようにデータセットを作成しているはずだが，
実際はデータセットには埋め込まれた偏りが存在する！

識別が難しい画像群

Antonio Torralba, Alexei A. Efros.
Unbiased Look at Dataset Bias.
CVPR, 2011.

Figure 3. Dataset Look-alikes: Above, ImageNet is trying to impersonate three different datasets. Here, the samples from ImageNet that are closest to the decision boundaries of the three datasets are displayed. Look-alikes using PASCAL VOC are shown below.

- SVMの識別面に近い画像群

切り出された車画像のみを対象とした Name That Dataset識別機

Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

- データセットには各々異なる目的で作られているので，その影響を除く
- バウンディングボックスが与えられている車の画像を対象
  - 5つのデータセット
    - PASCAL
    - ImageNet,
    - SUN09
    - LabelMe
    - Caltech101
- 結果：61%の識別率！

判別的な画像群

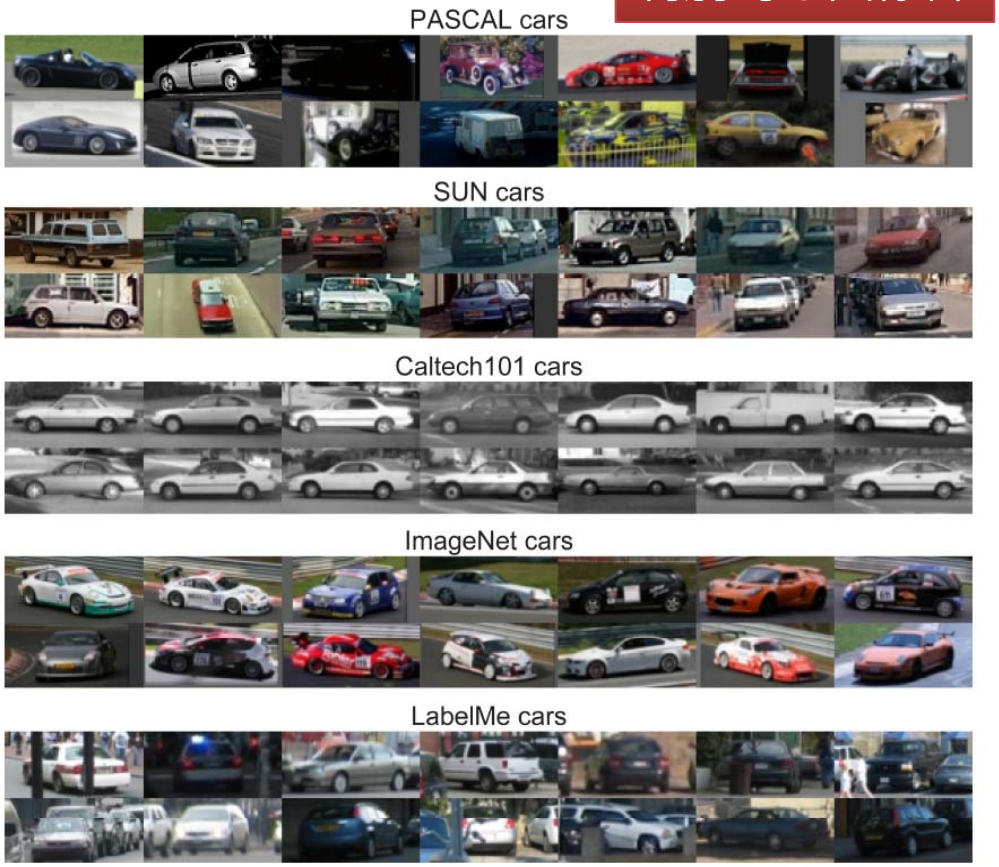


Figure 4. Most discriminative cars from 5 datasets

コンピュータビジョンのコミュニティはデータセットの偏りを取り除こうとしてきたがうまくいっていない.

The promise and perils of visual dataset 1/2

Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

- *Datasets have also played the leading role in making object recognition research look less like a black art and more like an experimental science.*
  - *データセットは物体認識研究を黒魔術ではなくより実験科学とする主役を演じてきた.*

- *Many people are worried that the field is now getting too obsessed with evaluation, spending more time staring at precision-recall curves than at pixels.*
  - *研究領域があまりに評価にとりつかれており，画像のピクセルを眺めるよりもprecision-recall曲線を眺めている時間が多い.*

- *There is concern that research is becoming too incremental, since a completely new approach will initially have a hard time competing against established, carefully fine-tuned methods.*
  - *研究があまりにインクリメンタルになりつつある懸念がある. なぜなら全く新しいアプローチははじめは確立されよくチューニングされた手法と争うには困難であるからである.*

The promise and perils of visual dataset 2/2

Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

- *Another concern is that our community gives too much value to "winning" a particular dataset competition, regardless of whether the improvement over other methods is statistically significant.*
  - *コンピュータビジョンコミュニティは，他の手法と比較して統計的に有意かどうかに関わらず特定のデータセットのコンペティションに勝つことに過度の価値を与えている.*

- *For PASCAL VOC, Everingham et al use the Friedman/Nemenyi test, which, for example, showed no statistically significant difference between the eight top-ranked algorithms in the 2010 competition.*
  - *2010年のPASCAL VOCにおいてトップランクの8つのアルゴリズム間に統計的有意差がない.*

本研究の主眼

- *There is a more fundamental question: are the datasets measuring the right thing, that is, the expected performance on some real-world task?*
  - *より本質的な質問：データセットは正しいものをはかっているのか，すなわち，ある実世界のタスクに対して期待されるパフォーマンスをはかっているのであろうか？*

The rise of the modern dataset

Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

データセットの発展：不公平と偏りに対する争いの物語

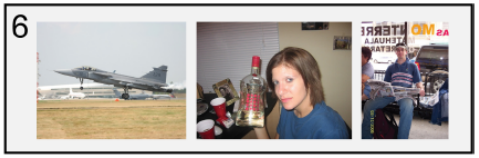
- COIL-100 dataset
  - 当時のモデルベースの考え方に対する反発
  - データドリブンなアピアランスモデルの採用

- 15 Scenes dataset, Corel Stock Photo
  - シンプルな背景への反発
  - 見た目の複雑さの採用

- Caltech101
  - Corelのようなプロフェッショナルが撮影した画像に対する反発（一部）
  - インターネット画像のwildnessの採用

- MSRC, LabelMe
  - 1つの物体が中心にあるというメンタリティへの反発
  - 多くの物体がある複雑なシーンの採用

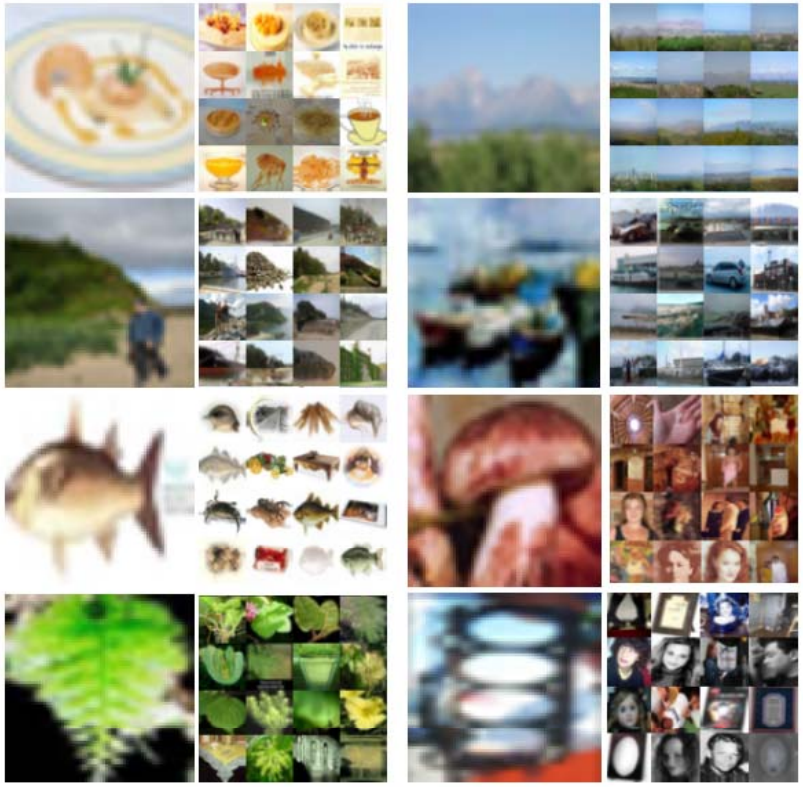
- PASCAL VOC
  - 以前のトレーニングとテスト基準への反発



- Tiny Images, ImageNet, SUN09
  - 実世界の複雑さに対して小さすぎるデータセットの学習とテストの不適さに対する反発

TinyImages

- A. Torralba, R. Fergus, W. T. Freeman. 80 million tiny images: a large dataset for non-parametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30(11), pp. 1958-1970, 2008.
- 8000万枚の画像データセット
- データが大量にあれば最近傍法のみで十分認識可能



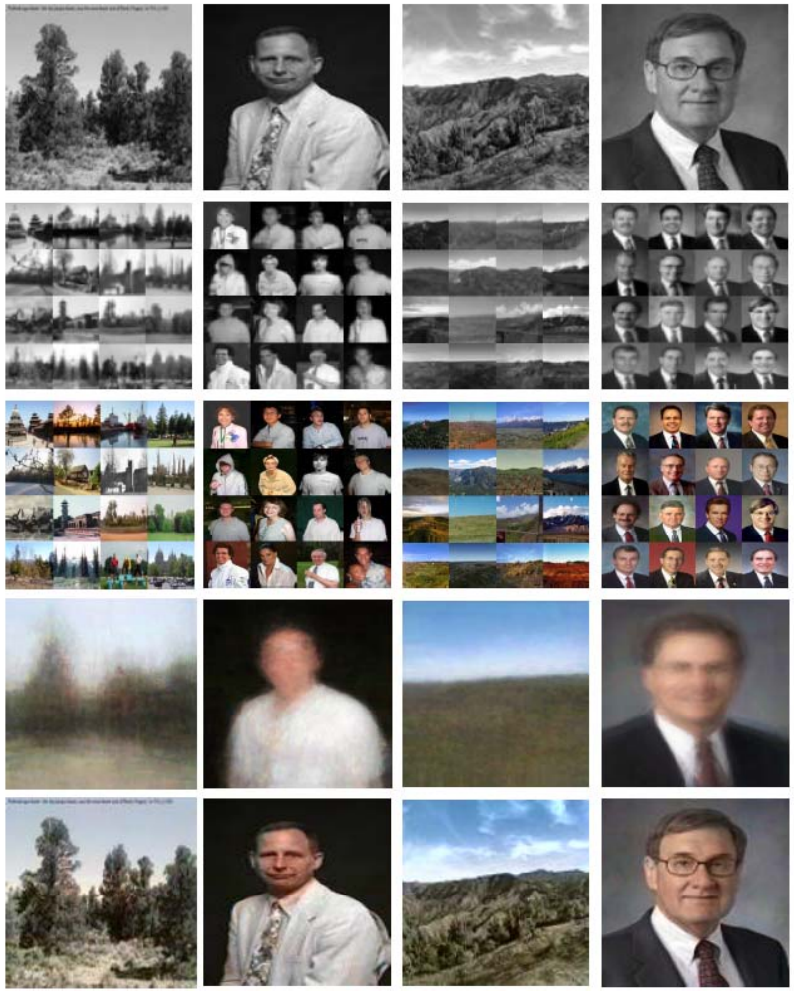


Fig. 1. 1st & 3rd columns: Eight 32 × 32 resolution color images. Despite their low resolution, it is still possible to recognize most of the objects and scenes. These are samples from a large dataset of 10^8 32 × 32 images we collected from the web which spans all visual object classes. 2nd & 4th columns: Collages showing the 16 nearest neighbors within the dataset to each image in the adjacent column. Note the consistency between the neighbors and the query image, having related objects in similar spatial arrangements. The power of the approach comes from the copious amount of data, rather than sophisticated matching methods.

ARISTA

- Xin-Jing Wang, Lei Zhang, Ming Liu, Yi Li, Wei-Ying Ma. ARISTA - Image Search to Annotation on Billions of Web Photos. In CVPR, 2010.
- 20億枚の画像データセットを利用した画像認識
- Near duplicated imageの活用．特定の名称まで認識可能.

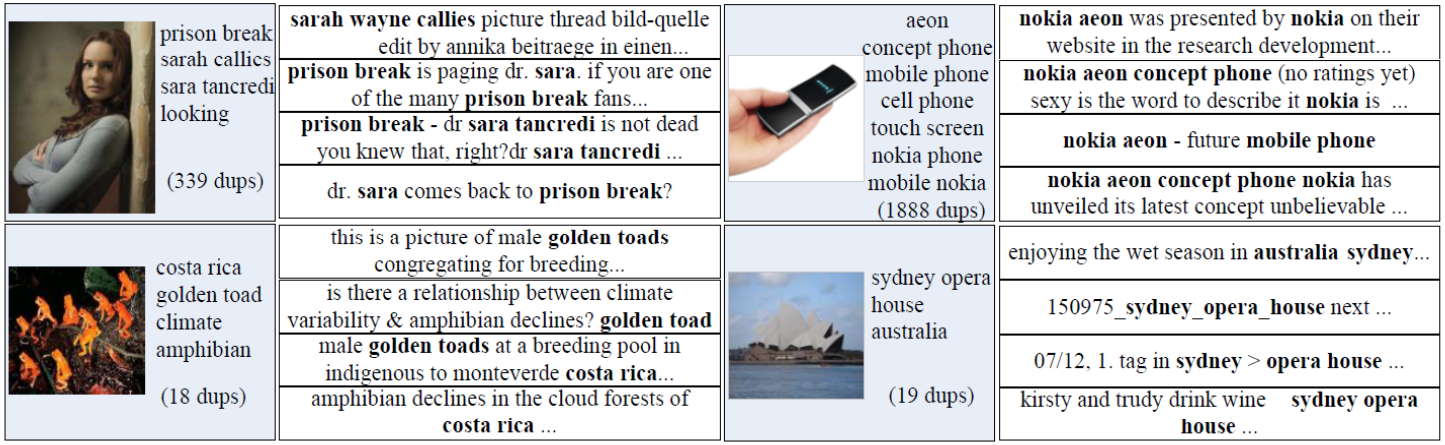


Figure 1. Examples showing that surrounding texts of near-duplicates have common terms which hit the semantics of a query image. The tags inside the image blocks are our annotation outputs. The common terms of each near-duplicate are highlighted in bold. Note that the detected tags are very specific. This is in contrast to most existing works that tend to generate general terms like sky, city, etc.

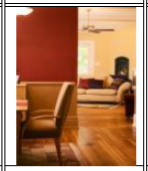
	2.4 M	80M	2B		2.4M	80M	2B
	(no results)	(no results)	*prison break*, **sarah callies**, **sara tancredi**, looking		(no results)	house paint, color	*house*, *paint*, wanta-toos, house painting, hardwood floor, interior design
	**michael jackson**	**michael jackson**, *rock pop*	**michael jackson**, sony music, *cd dvd*, *enter-tainment music*, *pop rock*		linu, *logo*	server, *software*, *logo*, credit card processing, *op-erating system*	**penguin**, *open source*, *virtual server*, *logo*, *operating system*
	**ipod touch**	**apple ipod**, *mp3 player*, **iphone**, wi fi, touch screen	**apple ipod**, *mp3 player*, wi fi, media player, **touch screen**, **mobile phone**		(no results)	(no results)	**bald eagle**, **haliaee-tus leucocephalus**, endangered species, fish wildlife, *eagle flight*

Figure 9. Annotation examples vs. dataset size. Bold-faced tags are perfect terms labeled by human subjects and italic ones are correct terms. Due to space limit, only the top five tags are shown. This figure suggests that larger dataset size ensures more accurate tags.

Measuring Dataset Bias

Antonio Torralba, Alexei A. Efros.
Unbiased Look at Dataset Bias.
CVPR, 2011.

- データセットを使って実世界を評価したい！
  - 実世界のデータセットを作る.
  - データセットに偏りが生じる可能性，，，

- 典型的な物体検出・識別機をあるデータセットで学習して他のデータセットでテストしたときにどれだけうまくいくかを評価する.
  - 検出：HOG + SVM
  - 識別：BoF + Gaussian Kernel +SVM
  - 対象：car + person

Measuring Dataset Bias結果

Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

データセットにどれだけ汎化性があるか

汎化性高い

簡単

データセットがどれだけ簡単か

task	Train on: \ Test on:	SUN09	LabelMe	PASCAL	ImageNet	Caltech101	MSRC	Self	Mean others	Percent drop
"car" classification	SUN09	**28.2**	29.5	16.3	14.6	16.9	21.9	28.2	19.8	**30%**
	LabelMe	14.7	**34.0**	16.7	22.9	43.6	24.5	34.0	24.5	**28%**
	PASCAL	10.1	25.5	**35.2**	43.9	44.2	39.4	35.2	32.6	**7%**
	ImageNet	11.4	29.6	36.0	**57.4**	52.3	42.7	57.4	34.4	**40%**
	Caltech101	7.5	31.1	19.5	33.1	**96.9**	42.1	96.9	26.7	**73%**
	MSRC	9.3	27.0	24.9	32.6	40.3	**68.4**	68.4	26.8	**61%**
	Mean others	10.6	28.5	22.7	29.4	39.4	34.1	53.4	27.5	48%
"car" detection	SUN09	**69.8**	50.7	42.2	42.6	54.7	69.4	69.8	51.9	**26%**
	LabelMe	61.8	**67.6**	40.8	38.5	53.4	67.0	67.6	52.3	**23%**
	PASCAL	55.8	55.2	**62.1**	56.8	54.2	74.8	62.1	59.4	**4%**
	ImageNet	43.9	31.8	46.9	**60.7**	59.3	67.8	60.7	49.9	**18%**
	Caltech101	20.2	18.8	11.0	31.4	**100**	29.3	100	22.2	**78%**
	MSRC	28.6	17.1	32.3	21.5	67.7	**74.3**	74.3	33.4	**55%**
	Mean others	42.0	34.7	34.6	38.2	57.9	61.7	72.4	44.8	**48%**
"person" classification	SUN09	**16.1**	11.8	14.0	7.9	6.8	23.5	16.1	12.8	**20%**
	LabelMe	11.0	**26.6**	7.5	6.3	8.4	24.3	26.6	11.5	**57%**
	PASCAL	11.9	11.1	**20.7**	13.6	48.3	50.5	20.7	27.1	**-31%**
	ImageNet	8.9	11.1	11.8	**20.7**	76.7	61.0	20.7	33.9	**-63%**
	Caltech101	7.6	11.8	17.3	22.5	**99.6**	65.8	99.6	25.0	**75%**
	MSRC	9.4	15.5	15.3	15.3	93.4	**78.4**	78.4	29.8	**62%**
	Mean others	9.8	12.3	13.2	13.1	46.7	45.0	43.7	23.4	**47%**
"person" detection	SUN09	**69.6**	56.8	37.9	45.7	52.1	72.7	69.6	53.0	**24%**
	LabelMe	58.9	**66.6**	38.4	43.1	57.9	68.9	66.6	53.4	**20%**
	PASCAL	56.0	55.6	**56.3**	55.6	56.8	74.8	56.3	59.8	**-6%**
	ImageNet	48.8	39.0	40.1	**59.6**	53.2	70.7	59.6	50.4	**15%**
	Caltech101	24.6	18.1	12.4	26.6	**100**	31.6	100	22.7	**77%**
	MSRC	33.8	18.2	30.9	20.8	69.5	**74.7**	74.7	34.6	**54%**
	Mean others	44.4	37.5	31.9	38.4	57.9	63.7	71.1	45.6	**36%**

Cross-dataset generalizationの結果

Antonio Torralba, Alexei A. Efros.
Unbiased Look at Dataset Bias.
CVPR, 2011.

- MSRCで学習

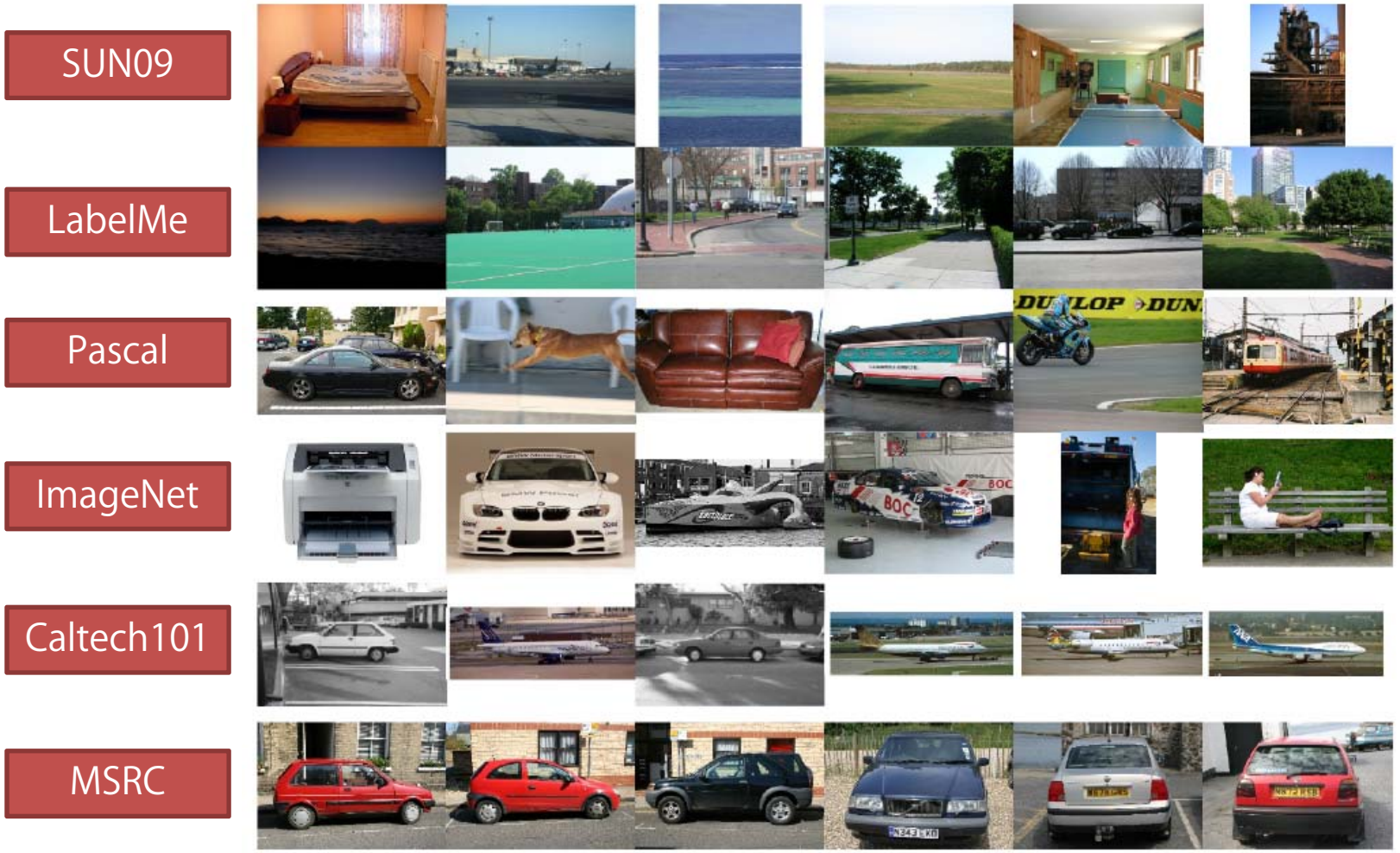


Figure 5. Cross-dataset generalization for "car" classification (full image) task, trained on MSRC and tested on (one per row): SUN, LabelMe, PASCAL, ImageNet, Caltech-101, and MSRC.

Measuring Dataset's Value

Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

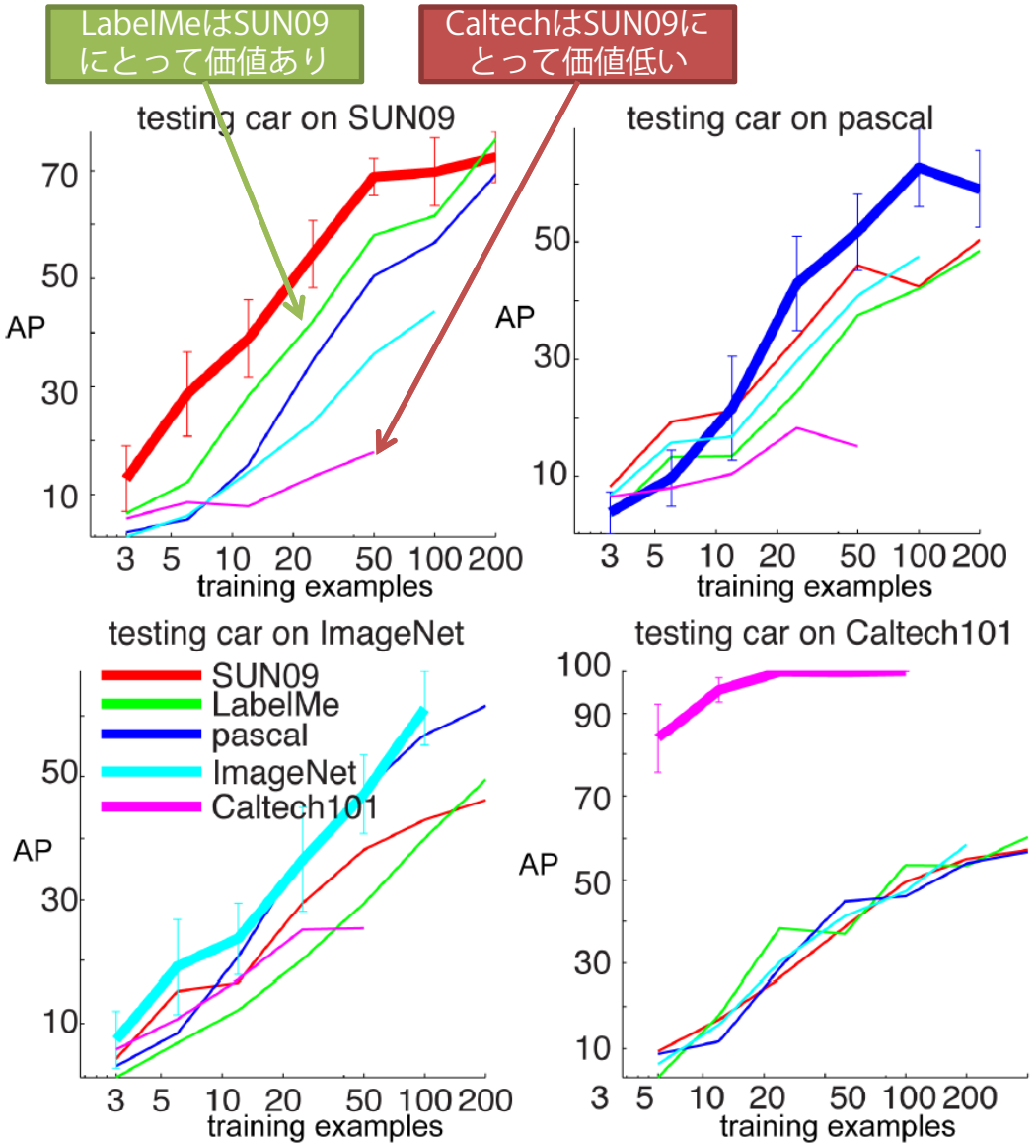


Figure 6. Cross-dataset generalization for "car" detection as function of training data

- 識別性能を上げたい！
  - 特徴量，識別則を改良
  - 識別性能向上させるためにはデータ数を増やす．

それほど簡単ではない☹

- 異なるドメインからデータを持ってきてデータ量を増やすことを考える
- 問題：あるデータセット訓練サンプルの価値に対して，他のデータセットのサンプルの相対的価値はどのくらいか？

Measuring Dataset's Valueの結果

Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

Table 3. "Market Value" for a "car" sample across datasets

	SUN09 market	LabelMe market	PASCAL market	ImageNet market	Caltech101 market
1 SUN09 is worth	1 SUN09	0.91 LabelMe	0.72 pascal	0.41 ImageNet	0 Caltech
1 LabelMe is worth	0.41 SUN09	1 LabelMe	0.26 pascal	0.31 ImageNet	0 Caltech
1 pascal is worth	0.29 SUN09	0.50 LabelMe	1 pascal	0.88 ImageNet	0 Caltech
1 ImageNet is worth	0.17 SUN09	0.24 LabelMe	0.40 pascal	1 ImageNet	0 Caltech
1 Caltech101 is worth	0.18 SUN09	0.23 LabelMe	0 pascal	0.28 ImageNet	1 Caltech
Basket of Currencies	0.41 SUN09	0.58 LabelMe	0.48 pascal	0.58 ImageNet	0.20 Caltech

LabelMeの1サンプルは，PASCALベンチマークにおいて
PASCALの0.26サンプル分に相当する！

- 問題
  - 1250のPASCALサンプルで学習済みの識別機がある．この識別機をPASCALデータセットにおいてAPを10%性能向上させたい．何枚のLabelMeデータセットが必要か？
- 答え
  - 1/0.26 x 1250 x 10 =50,000 LabelMe samples!

Fig.6のグラフから関係式を得る．(1-AP)=n^bとしてフィッティングするとだいたいb=-1/5となる．
n=1250のときのAPに0.1（10%）増加させたときのnはだいたい18000となる．オーダーとして1250の10倍.

データセットはどう作るべきか？

Antonio Torralba, Alexei A. Efros. Unbiased Look at Dataset Bias. CVPR, 2011.

- Selection Bias
  - データセットが特定の種類の画像を好むために生じる偏り.
    - E.g. Street scenes, nature scenes, images via internet search
  - 対策
    - 手動でなく自動で収集
    - ラベルのないデータを収集して，クラウドソーシングでラベル付け

- Capture Bias
  - 特定の構図を好むために生じる偏り
    - Mugで画像検索すると取っ手が右側にあるのが多い.
  - 対策
    - 画像にいろんな変換を加える.

- Negative Set Bias
  - 対象とするクラス以外のクラスは無限に存在するにも関わらず その他のクラスを少数のクラスで代表させることで生じる偏り
  - 対策
    - 他のデータセットからネガティブを集める
    - ラベルなしのデータセットから一般的な識別機で困難なネガティブを 収集する．そのデータから手動でtrue positiveを除く.

原文を読んでみよう！

Antonio Torralba, Alexei A. Efros.
Unbiased Look at Dataset Bias.
CVPR, 2011.

from multiple countries [3]) can somewhat decrease selection bias. However, it might be even better to start with a large collection of unannotated images and label them by crowd-sourcing.

Capture Bias: Professional photographs as well as pho-
tos collected using key
erably from the captur
is that the object is alr
age. Searching for "n
reveal another kind of
a right-facing handle.
gies, one way to deal
transformations to redu
left-right [8, 9] (but not
way), or jittering the in
formations [18]. Anoth
ating various automatic

Negative Set Bias: As we have shown, having a rich and unbiased negative set is important to classifier performance. Therefore, datasets that only collect the things they are interested in might be at a disadvantage, because they are not modeling the rest of the visual world. One remedy, proposed in this paper, is to add negatives from other datasets. Another approach, suggested by Mark Everingham, is to use a few standard algorithms (e.g. bag of words) to actively mine hard negatives as part of dataset construction from a very large unlabelled set, and then manually going through them to weed out true positives. The down side is that the resulting dataset will be biased against existing algorithms.

This paper is only the start of an important conversation about datasets. We suspect that, despite the title, our own biases have probably crept into these pages, so there is clearly much more to be done. All that we hope is that our work will start a dialogue about this very important and underappreciated issue.

Acknowledgements: The authors would like to thank the Eyjafjallajokull volcano as well as the wonderful *kirs* at the Buvette in Jardin du Luxembourg for the motivation (former) and the inspiration (latter) to write this paper. This work is part of a larger effort, joint with David Forsyth and Jay Yagnik, on understanding the benefits and pitfalls of using large data in vision. The paper was co-sponsored by ONR MURIs N000141010933 and N000141010934.

Disclaimer: No graduate students were harmed in the production of this paper. Authors are listed in order of increasing procrastination ability.

References

[1] N. Dalal and B. Triggs. Histogram of oriented gradients for human detection. 2005. 1521, 1524

[2] D. DeCoste and M. Burl. Distortion-invariant recognition via jittered queries. In *CVPR*, 2000. 1528

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. 2009. 1523, 1524, 1528

[4] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *CVPR*, 2009. 1523, 1524

[5] L. Duan, I. W.-H. Tsang, D. Xu, and S. J. Maybank. Domain transfer svm for video concept detection. In *CVPR*, 2009. 1524

egory dataset. Technical report, California Institute of Technology, 2007. 1528

[10] J. Hutchison. Culture, communication, and an informati age madonna. In *IEEE Professional Communication Society Newsletter*, volume 45, 2001. 1523

[11] T. Malisiewicz and A. A. Efros. Recognition by association via learning per-exemplar distances. In *CVPR*, 2008. 1525

[12] S. A. Nene, S. K. Nayar, and H. Murase. Columbia object image library (coil-100). Technical Report CUCS-006-96, Columbia Univ., 1996. 1523

[13] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal in Computer Vision*, 42:145–175, 2001. 1521, 1523

[14] J. Ponce, T. L. Berg, M. Everingham, S. Lazebnik, M. Marszałek, C. Schmid, B. Russell, A. Torralba, C. Williams, J. Zhang, and A. Zisserman. Dataset issues in object recognition. In *Toward Category-Level Object Recognition*. Springer, 2006.

[15] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[16] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: a database and web-based tool for image annotation. 77(1-3):157–173, 2008.

[17] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *ECCV*, 2010. 1524

[18] A. Torralba, R. Fergus, and W. T. Freeman. 80 million tiny images: a large database for non-parametric object and scene recognition. *IEEE PAMI*, 30(11):1958–1970, November 2008. 1521, 1523, 1528

[19] J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *ICCV*, 2005. 1523, 1524

[20] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR*, 2010. 1523, 1524

[21] J. Yang, R. Yan, and A. G. Hauptmann. Cross-domain video concept detection using adaptive svms. MULTIMEDIA '07, 2007. 1524

Disclaimer: No graduate students were harmed in the production of this paper. Authors are listed in order of increasing procrastination ability.

免責事項：本論文の作成にあたり，いずれの大学院生も被害を被っていない．著者はぐずぐずする能力が増える順に並んでいる．

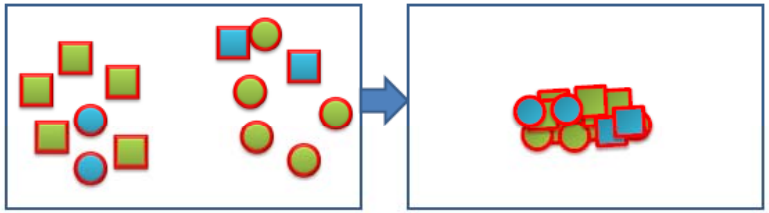
ドメイン適応（Domain adaptation）

- Brian Kulis, Kate Saenko, and Trevor Darrell. What You Saw is Not What You Get: Domain Adaptation Using Asymmetric Kernel Transforms. CVPR, 2011.

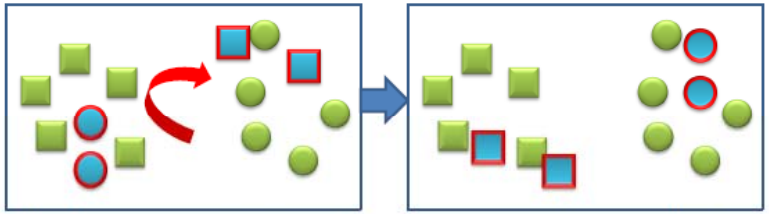
Figure 1. We address the problem of adapting object models trained on a particular source dataset, or domain (left), to a target domain (right).

ドメイン適応問題と論文のキーアイデア

B. Kulis, K. Saenko, and T. Darrell. What You Saw is Not What You Get: Domain Adaptation Using Asymmetric Kernel Transforms. CVPR, 2011.



(a) A symmetric transformation – the same rotation and scaling applied to both domains (green and blue) – cannot separate classes (circles and squares)



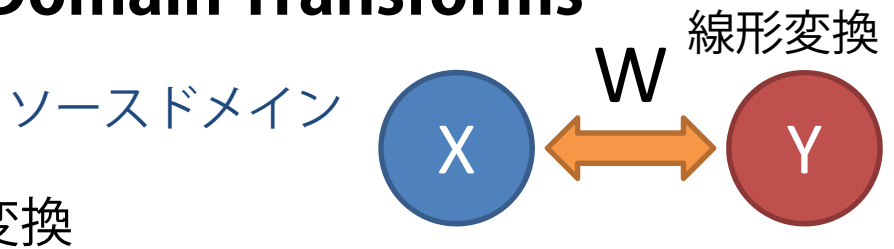
(b) An asymmetric transformation – a rotation applied only to blue domain – successfully compensates for domain shift

Figure 2. A conceptual illustration of how an asymmetric domain transform (this paper) can be more flexible than a symmetric one [19].

- ドメイン適応問題
  - ソースドメインと異なる特徴分布をもつターゲットドメインが与えられたとき，ソースドメインで学習したモデルをテスト時にどのように効率的に扱えばよいのか？
- 本論文でのキーアイデア
  - 双方のドメインからの教師付データを用いた，あるドメインからもう一方のドメインに点群を写像する非対称の非線形変換の学習

Domain Adaptation Using Regularized Cross-Domain Transforms

B. Kulis, K. Saenko, and T. Darrell. What You Saw is Not What You Get: Domain Adaptation Using Asymmetric Kernel Transforms. CVPR, 2011.



- 非対称変換

損失関数

目的関数

$$\min_W r(W) + \lambda \sum_i c_i(X^T W Y), \quad (1)$$

内積だから大きくしたい

$r(W) = \frac{1}{2}\|W\|_F^2$ 正則化項

任意の行列が利用可能！

xとyが同じクラス → $c_i(X^T W Y) = (\max(0, \ell - \boldsymbol{x}^T W \boldsymbol{y}))^2$

xとyが異なるクラス → $c_i(X^T W Y) = (\max(0, \boldsymbol{x}^T W \boldsymbol{y} - u))^2$

内積だから小さくしたい

Asymmetric Regularized Cross-domain transformation problem with similarity and dissimilarity constraints (ARC-t)

- カーネル化 (see Appendix)

RBFカーネル等に置換可能！

$$K_A = X^T X \quad K_B = Y^T Y$$

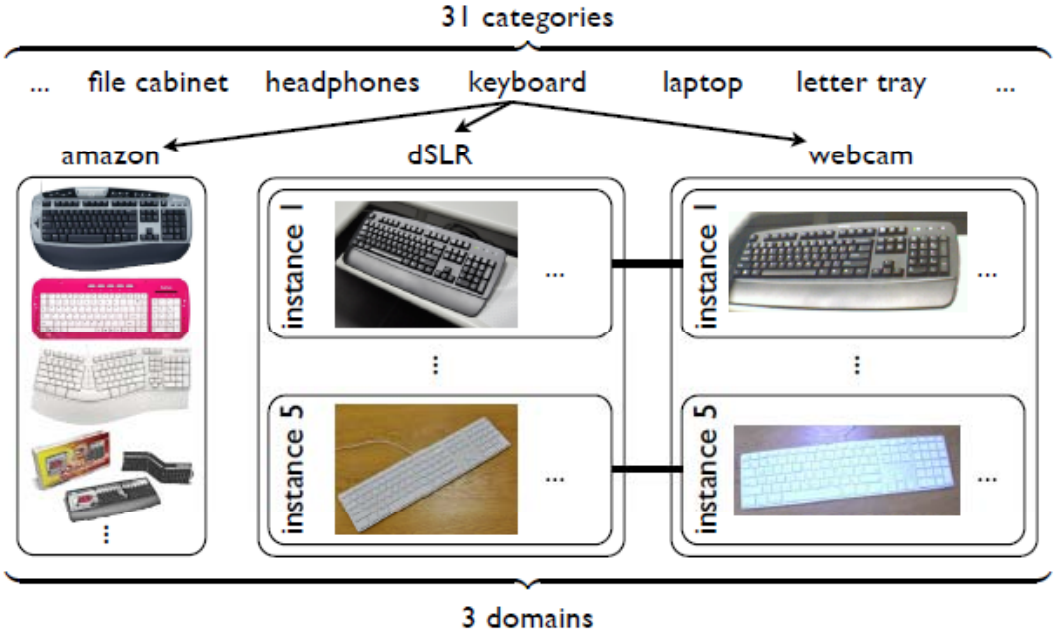
$$\min_L r(L) + \lambda \sum_i c_i(K_A^{1/2} L K_B^{1/2}) \quad (2)$$

$$W = X K_A^{-1/2} L K_B^{-1/2} Y^T$$

式（1）をWに関して解くことと，式（2）をLに関して解くことは等価.

正則化項は凸なので多種の最小化の最適化手法を利用可能.

実験



B. Kulis, K. Saenko, and T. Darrell. What You Saw is Not What You Get: Domain Adaptation Using Asymmetric Kernel Transforms. CVPR, 2011.

Domain adaptation dataset

・31カテゴリ

・3ドメイン，amazon，一眼レフ，webcam

- Same category experiment
  - 全てのクラスの対応関係が学習可能
  - 識別機：ARC-t + 最近傍則
  - 訓練：ソース20サンプル，ターゲット3サンプル
  - 比較手法にはあらかじめKCCAを適用
    - ドメイン間の次元の違いを吸収するため

- New category experiment
  - 訓練時，一部（半分）のカテゴリの対応関係しか学習できない
  - ソース20サンプル，ターゲット10サンプル
  - テスト時，ターゲットドメインでは新規のクラスで，ソースドメインではラベル付きサンプルがある．
  - ターゲットドメインのデータを学習した変換でソースドメインに写像する．そしてソースドメインで最近傍則を適応する．

実験結果

B. Kulis, K. Saenko, and T. Darrell. What You Saw is Not What You Get: Domain Adaptation Using Asymmetric Kernel Transforms. CVPR, 2011.

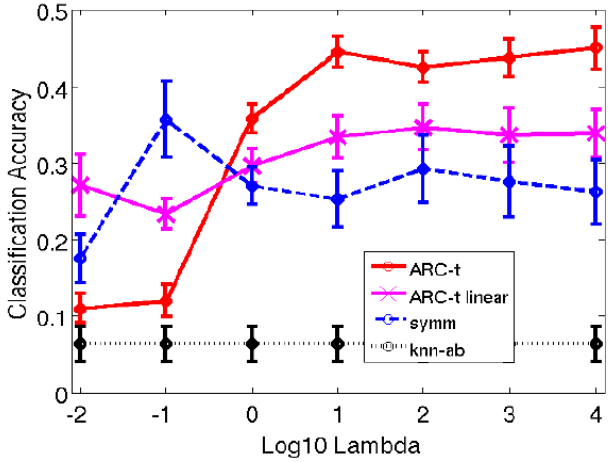


Figure 5. Plot of classification accuracy as a function of the learning rate lambda over the webcam800-dslr600 new categories experiment. This plot also shows a comparison between learning a linear transformation (ARC-t linear) and a non-linear transformation (ARC-t).

Figure 4. Examples of the 5 nearest neighbors retrieved for a dslr query image (left image) from the amazon dataset, using the non-adapted **knn-ab** baseline in Table 1 (top row of smaller images) and the learned cross-domain **ARC-t** kernel (bottom row).

		Baselines / Existing Methods		This Paper	
Domain A	Domain B	**knn-ab**	**symm** [19]	**ARC-t**	**ARC-t linear**
webcam	dslr	8.4	30.3	**37.4**	32.5
webcam-800	dslr-600	9.7	35.8	**45.0**	34.8
webcam-surf	dslr-sift	9.7	17.0	**24.8**	20.6

アトリビュート（Attribute）

- 物体カテゴリ間で共有される人間が理解可能な属性

- アトリビュートの主な適応先の分類
  1. 一般もしくは見慣れない物体の記述
  2. Zero-shot認識，知識転移，転移学習
  3. 物体識別を補助する中間特徴

S. J. Hwang, F. Sha, and K. Grauman. Sharing Features Between Objects and Their Attributes. CVPR, 2011.

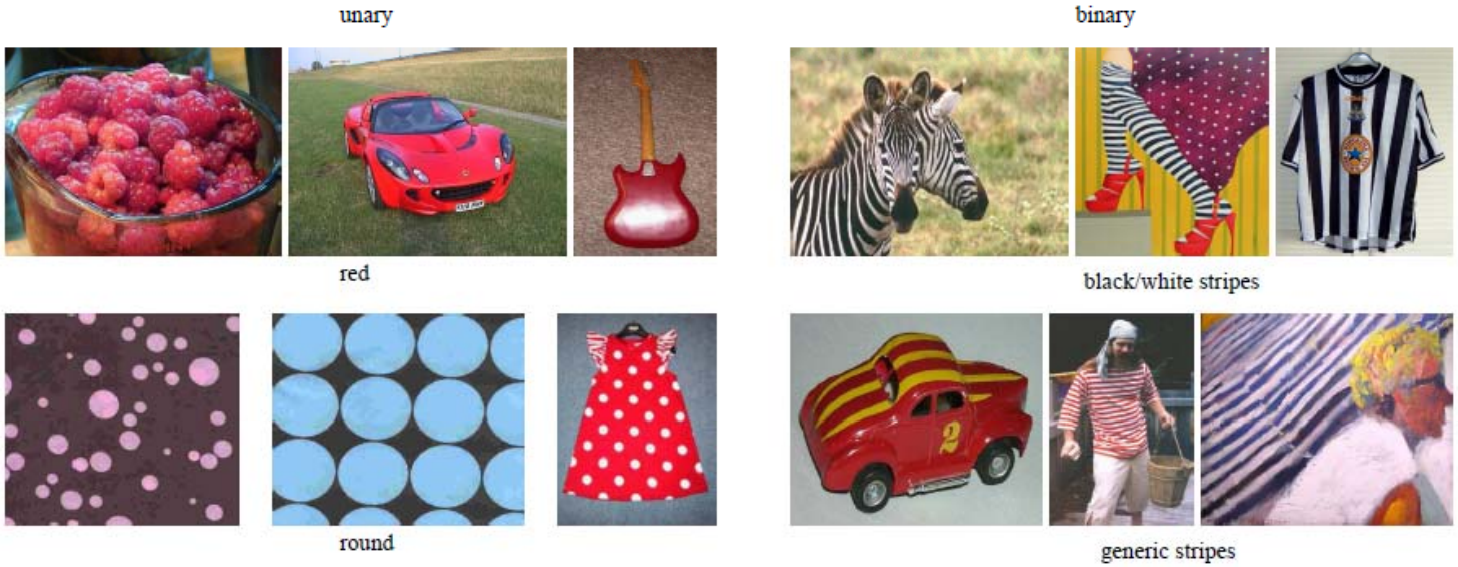
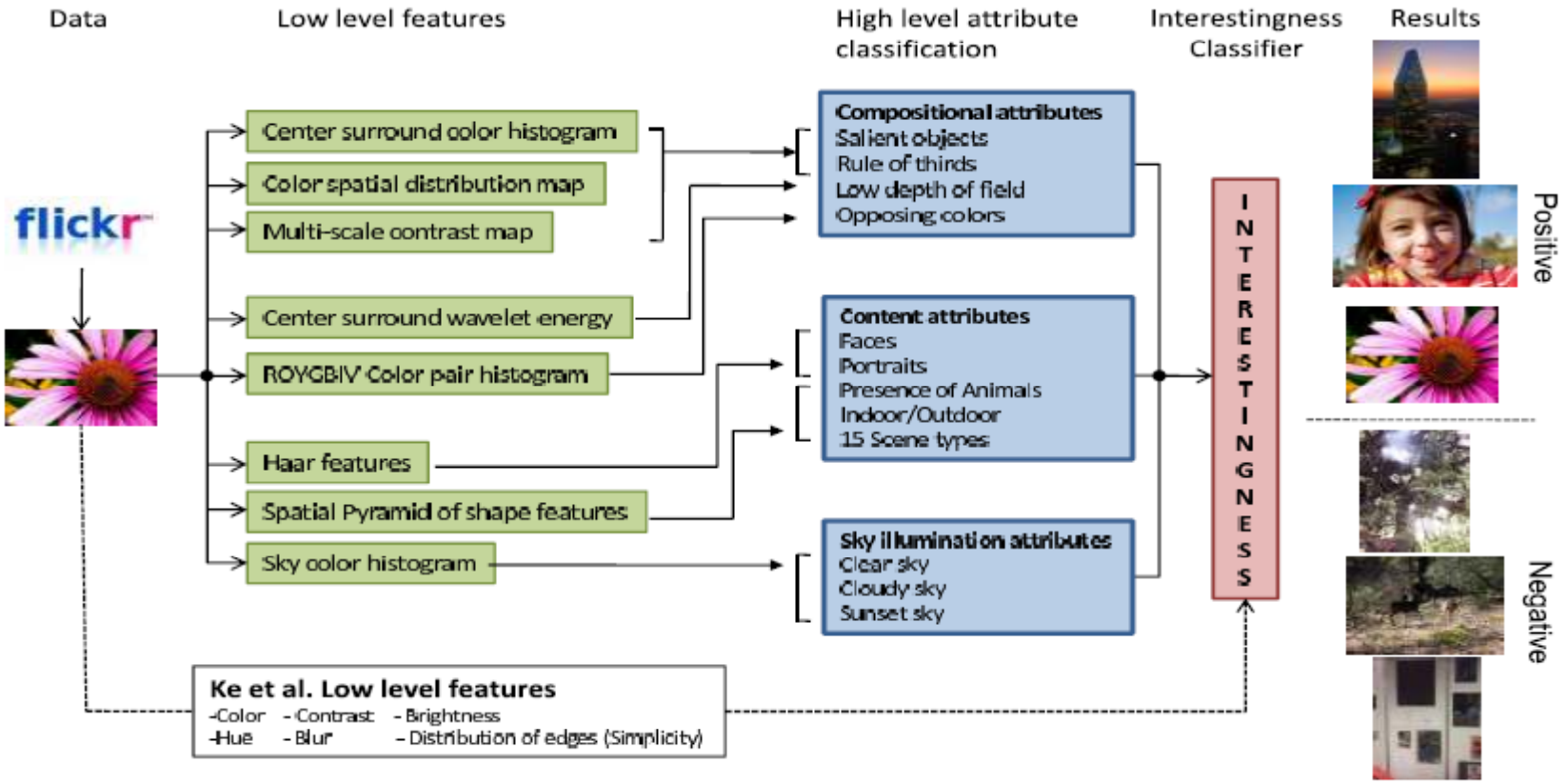


Figure 1: *Examples of different kinds of attributes. On the left we show two simple attributes, whose characteristic properties are captured by individual image segments (appearance for red, shape for round). On the right we show more complex attributes, whose basic element is a pair of segments.*

V. Ferrari and A. Zisserman. Learning visual attributes. In NIPS, 2008.

アトリビュートの美と魅力予測への応用

- Sagnik Dhar Vicente Ordonez Tamara L Berg. High Level Describable Attributes for Predicting Aesthetics and Interestingness. CVPR, 2011.



データセット

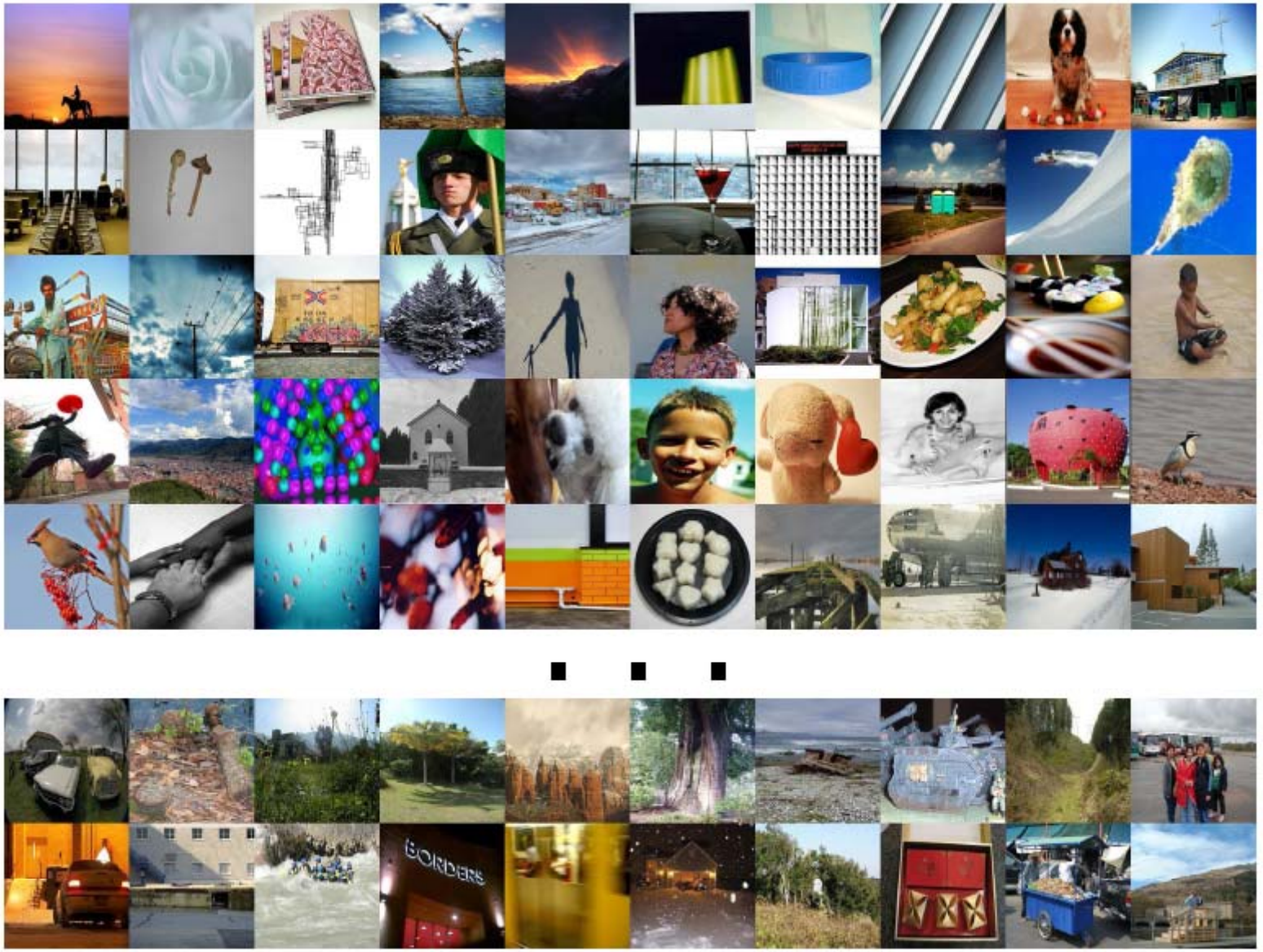
S. Dhar, V. Ordonez, and T. L Berg. High Level Describable Attributes for Predicting Aesthetics and Interestingness. CVPR, 2011.

- 美（Aesthetics）
  - DPChallenge website
  - http://www.dpchallenge.com/
  - 人手によるratingあり

- 魅力（Interestingness）
  - Flickr's "interestingness" measure
  - http://www.flickr.com/explore/interesting/
  - *There are lots of elements that make something 'interesting' (or not) on Flickr. Where the clickthroughs are coming from; who comments on it and when; who marks it as a favorite; its tags and many more things which are constantly changing. Interestingness changes over time, as more and more fantastic content and stories are added to Flickr.*
  - http://www.barcinski-jeanjean.com/entries/endlessintrestingness/

実験結果（interestingness）

S. Dhar, V. Ordonez, and T. L Berg. High Level Describable Attributes for Predicting Aesthetics and Interestingness. CVPR, 2011.



アトリビュートの画像検索への応用

- Matthijs Douze, Arnau Ramisa, and Cordelia Schmid. Combining attributes and Fisher vectors for efficient image retrieval. CVPR, 2011.

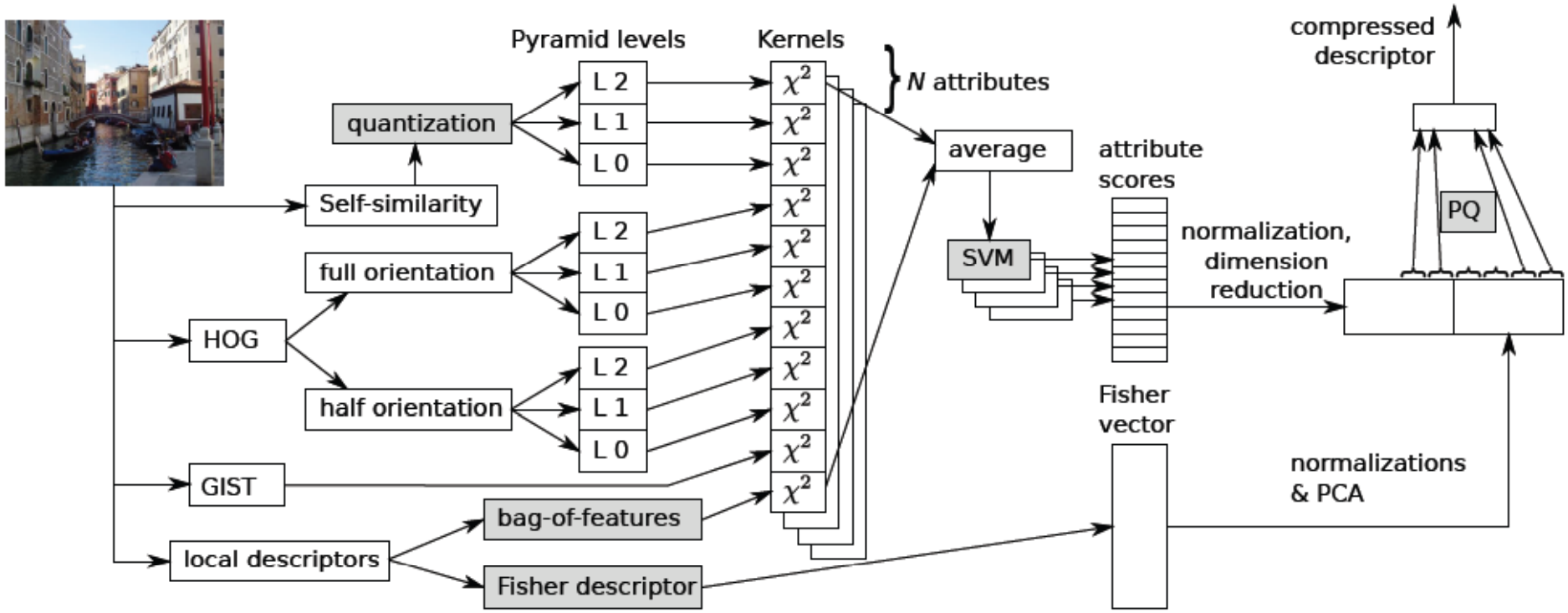


Figure 1. Computation of the attribute + Fisher descriptors for an image. Steps represented in gray require a learning stage.

2659属性
L. Torresani, M. Summer, and A. Fitzgibbon. Efficient object category recognition using classemes. In ECCV, 2010.

実験結果

M. Douze, A. Ramisa, and C. Schmid. Combining attributes and Fisher vectors for efficient image retrieval. CVPR, 2011.

Figure 3. Comparison of the retrieval results obtained with the Fisher vector, the attribute features, and their combination. The top row shows the query image, the remaining rows the first three retrieved images for the different descriptors.

Descriptor	dimension	mAP
BOF k=1000 [6]	1000	41.1
Fisher k=64 [17]	4096	≈ 60
Fisher k=4096 [17]	262144	70.5
VLAD k=64 [8]	8192	52.6
Fisher (F), k=64, L2 dist.	4096	59.5
Attributes (A), L2 dist.	2659	55.0
A + F, F-weight $\times 1$	6755	64.5
A + F, F-weight $\times 2$	6755	69.5
A + F, F-weight $\times 2.3$	6755	69.9

Table 1. Comparison of the different descriptors and their combination on the Holidays dataset.

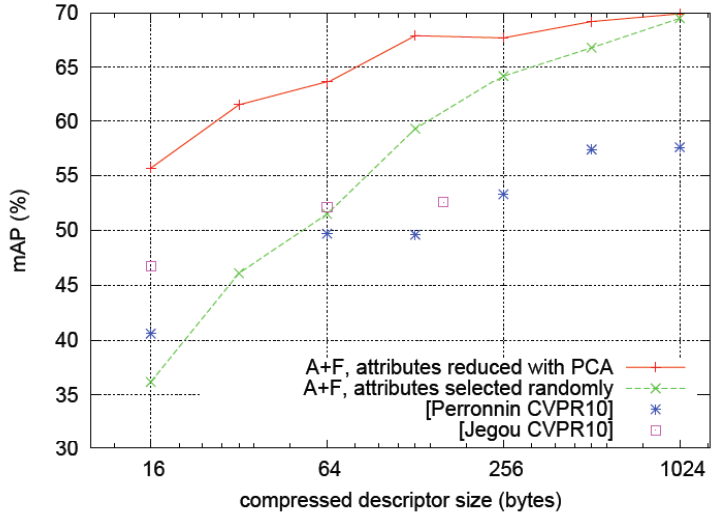


Figure 5. Performance of the A+F descriptor after dimension reduction and descriptor encoding on the Holidays dataset. The Fisher vectors are always reduced with PCA.

マルチタスク学習（Multi-task learning）

- Sung Ju Hwang, Fei Sha, and Kristen Grauman. Sharing Features Between Objects and Their Attributes. CVPR, 2011.

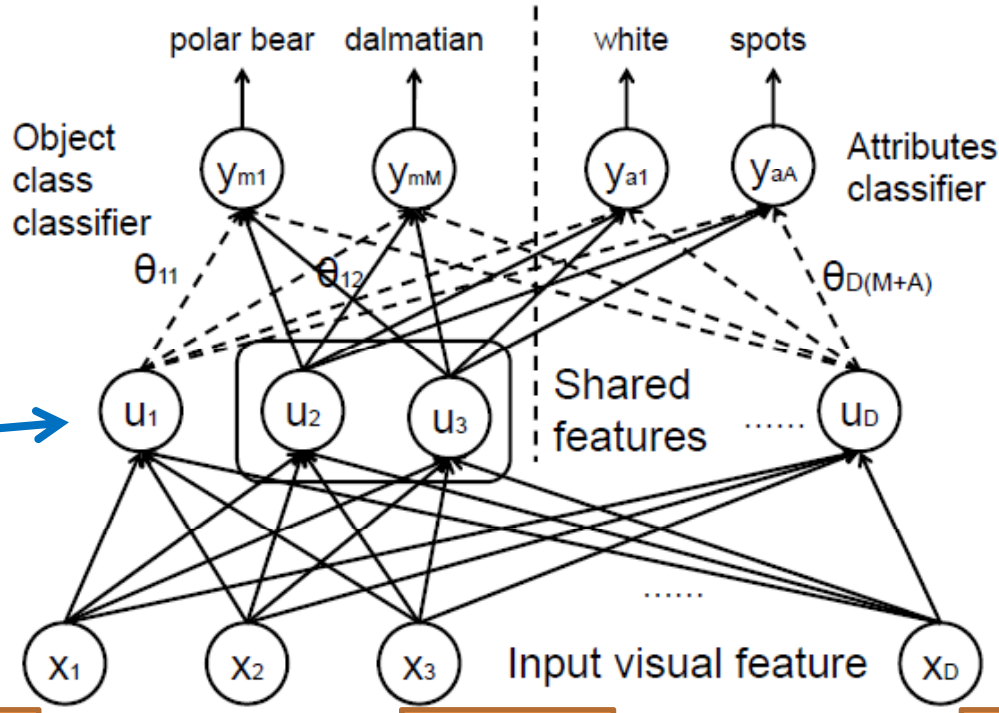
- 従来のアトリビュートを用いた物体認識
  - 「物体・アトリビュート間学習」と「アトリビュート・画像特徴間学習」が独立

- 提案するアトリビュートを用いた物体認識
  - 「物体・アトリビュート間学習」と「アトリビュート・画像特徴間学習」が共有特徴を通じてお互いに影響を与える.

- 仮定
  - 双方の予測タスクはオリジナルの画像特徴空間においてある共有する構造に依存している.

双方の識別モデルに従う共有した低次元表現を学習するアプローチの提案

モデル



S. J. Hwang, F. Sha, and K. Grauman. Sharing Features Between Objects and Their Attributes. CVPR, 2011.

なるべく低次元の共有特徴を求めたい

クラス・アトリビュートに関して

損失関数

ラベル

正則化項

$$\Theta^*, U^* = \arg\min \sum_t \sum_n \ell(\theta_t^{\mathrm{T}} U^{\mathrm{T}} x_n, y_{nt}) + \gamma \|\Theta\|_{2,1}^2$$

$$\Theta \in \mathbb{R}^{D \times T}$$

共有特徴からクラスへの重み

画像特徴から共有特徴への変換

画像特徴

なるべく低次元の共有特徴を求める工夫はここ！

正則化項

$$\|\Theta\|_{2,1} = \sum_{d=1}^{D} \sqrt{\sum_t \theta_{td}^2}$$

全てのtに関して0となる時だけ，この項が0となる

最小の非ゼロ行を選択
つまり，
最少の共有特徴しか残らない

計算の工夫

正則化項がスムーズでないので最適化問題は困難！

S. J. Hwang, F. Sha, and K. Grauman. Sharing Features Between Objects and Their Attributes. CVPR, 2011.

$$\Theta^*,\ U^* = \arg\min \sum_t \sum_n \ell(\theta_t^{\mathrm{T}} U^{\mathrm{T}} x_n, y_{nt}) + \gamma \|\Theta\|_{2,1}^2$$

式変形

M. P. A. Argyriou, T. Evgeniou. Convex Multi-task Feature Learning. Machine Learning, 73(3):243–272, 2008.

$$W^*,\ \Omega^* = \arg\min \sum_t \sum_n \ell(w_t^{\mathrm{T}} x_n, y_{nt})$$

安定化

$$+ \gamma \sum_t w_t^{\mathrm{T}} \Omega^{-1} w_t + \gamma\epsilon\, \mathrm{Trace}(\Omega^{-1}),$$

$$w_t = U\theta_t$$

$$\Omega^* = U^*\, \mathrm{Diag}\left(\left\{ \frac{\|\Theta_d\|_2}{\|\Theta\|_{2,1}} \right\}_{d=1}^{\mathrm{D}} \right) U^{*\mathrm{T}}$$

SVMと同じ！

wtとΩを交互に最適化すればよい

$$\hat{w}_t^* = \arg\min \sum_n \ell(\hat{w}_t^{\mathrm{T}} z_n, y_{nt}) + \gamma \|\hat{w}_t\|_2^2,$$

$$z_n \leftarrow \Omega^{1/2} x_n, \quad \hat{w}_t \leftarrow \Omega^{-1/2} w_t.$$

$$\Leftrightarrow \quad \Omega = \frac{(WW^{\mathrm{T}} + \epsilon I)^{1/2}}{\mathrm{Trace}\left[(WW^{\mathrm{T}} + \epsilon I)^{1/2}\right]}.$$

物体とアトリビュートの正則化項の分離

$$W^*,\ \Omega^* = \arg\min \sum_t \sum_n \ell(w_t^{\mathrm{T}} x_n, y_{nt}) + \epsilon\, \mathrm{Trace}(\Omega^{-1}) + \sum_{t=1}^{\mathrm{M}} \gamma_M w_t^{\mathrm{T}} \Omega^{-1} w_t + \sum_{t=\mathrm{M}+1}^{\mathrm{T}} \gamma_A w_t^{\mathrm{T}} \Omega^{-1} w_t$$

物体

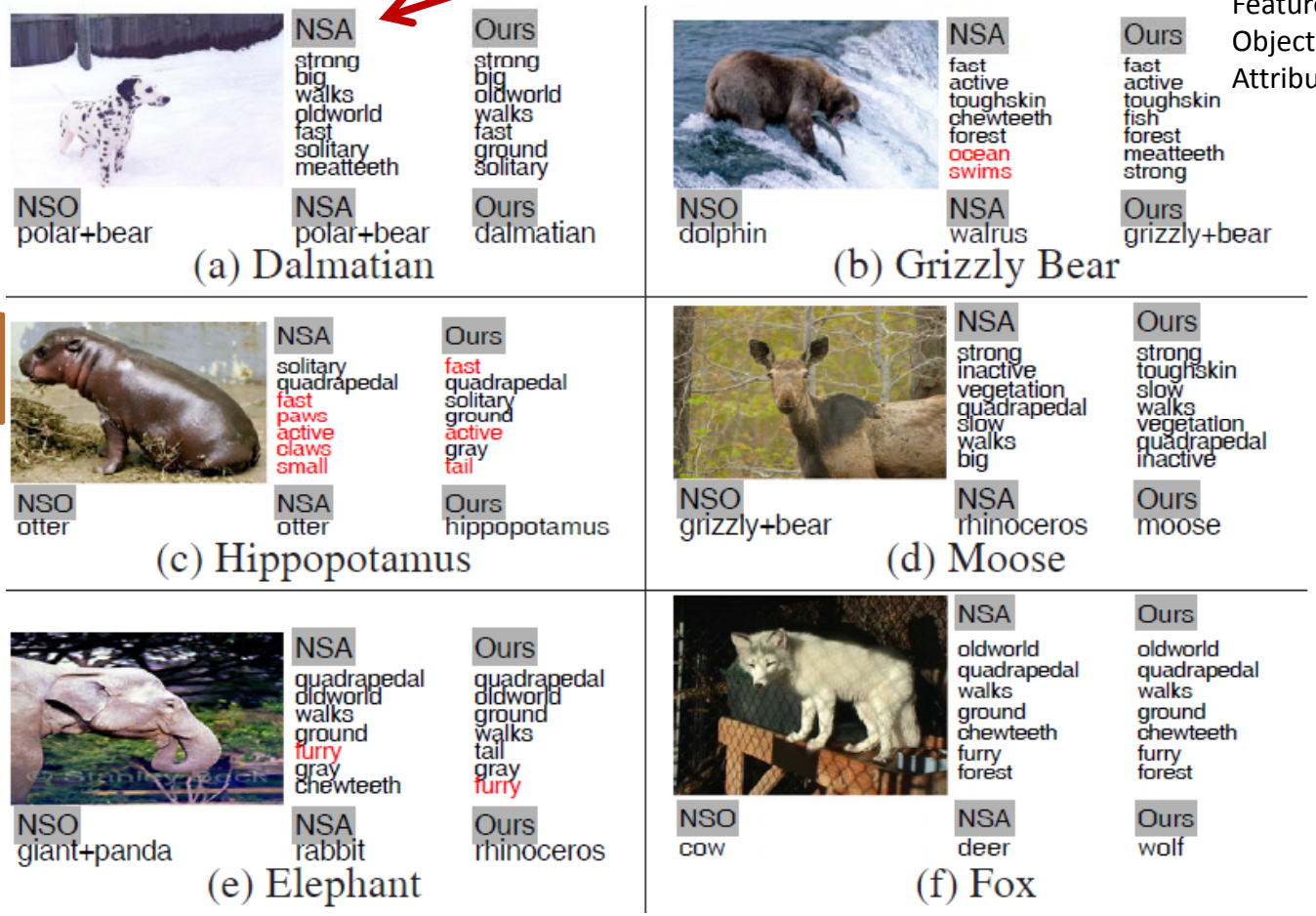
アトリビュート

実験結果

アトリビュート＋ロジスティック回帰

特徴+カイ2乗
カーネル+SVM

S. J. Hwang, F. Sha, and K. Grauman. Sharing Features Between Objects and Their Attributes. CVPR, 2011.



	50-class Animals Dataset				8-class Scenes Dataset			
Method / % train data	10%	20%	40%	60%	10%	20%	40%	60%
No sharing-Obj. (NSO)	31.96	38.12	44.08	48.03	76.76	79.75	83.03	83.74
No sharing-Attr. (NSA)	31.03	35.61	41.12	43.59	57.77	58.98	60.50	60.78
Sharing-Obj. (Ours)	37.08	41.01	46.46	49.15	78.76	81.49	85.05	86.06
Sharing+Attr. (Ours)	36.73	42.60	47.70	50.94	78.09	81.62	85.89	87.01
% gain over NSO	**14.92%**	**11.75%**	**8.21%**	**6.06%**	**1.73%**	**2.34%**	**3.44%**	**3.90%**
% gain over NSA	**18.37%**	**19.63%**	**16.00%**	**16.86%**	**35.17%**	**38.39%**	**41.97%**	**43.16%**

知識転移（Knowledge transfer）

- Marcus Rohrbach, Michael Stark, and Bernt Schiele. Evaluating Knowledge Transfer and Zero-Shot Learning in a Large-Scale Setting. CVPR, 2011.

- 目的
  - 近年提案されている知識転移のアプローチを再検討し，大規模データで評価を行うこと

- 知識転移のアプローチを3つに分類
  1. クラス空間における階層構造の利用
  2. 物体を表現するためにアトリビュートの利用
  3. 物体間の直接的な類似度の利用

- 知識転移の問題設定を2つに分類
  - 知識共有
    - 全てのクラス間で知識を共有する
    - 良い識別性能を期待
  - ゼロショット認識
    - 見たことのない物体クラスを認識する

知識転移のアプローチ

M. Rohrbach, M. Stark, and B. Schiele. Evaluating Knowledge Transfer and Zero-Shot Learning in a Large-Scale Setting. CVPR, 2011.

- 階層構造の利用
  - Inner WordNet nodes model
  - All WordNet nodes model
  - Leaf nodes, cost sensitive

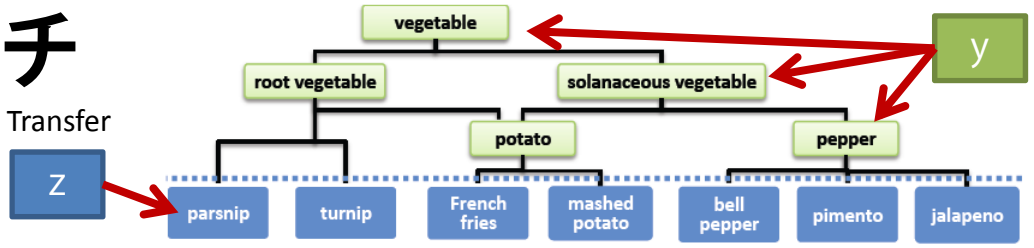


Figure 1: ISVLRC10 subgraph. Leaf (blue), inner nodes (green).

$$s^{inn}(z_l|x) = \frac{\sum_{y_i \in H_{z_l}} s(y_i|x)}{|H_{z_l}|}$$

内ノードの識別機

親ノードの集合

葉ノードの識別機

$$s^{all}(z_l|x) = \frac{s(z_l|x) + \sum_{y_i \in H_{z_l}} s(y_i|x)}{1 + |H_{z_l}|}$$

$$s^{cost}(z_l|x) = -\sum_{z_i \in Z_l} c^{z_l}_{z_i} p(z_i|x)$$

肩にかかっているけど積では???

階層コスト

アトリビュート識別機

$$s^{attr}(z_l|x) = \frac{\sum_{m=1}^{M} s(a_m|x)^{a_m^{z_l}}}{\sum_{m=1}^{M} a_m^{z_l}}$$

アトリビュートと葉ノードの関係を示す指標

前提：物体とアトリビュートの関係はgiven

- アトリビュートの利用

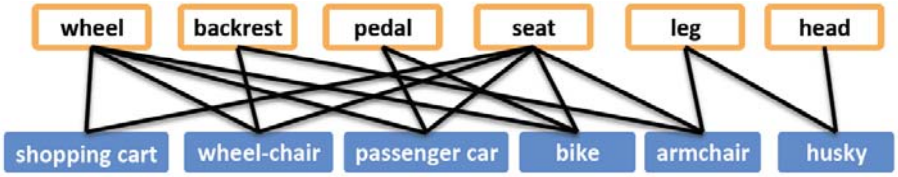


Figure 2: Example part attributes (orange), object classes (blue).

- 直接的類似度の利用

前提：物体間の関係はgiven

$$s^{dir}(z_l|x) = \frac{\sum_{k=1}^{K} s(z_k|x)}{K}$$

実験結果

M. Rohrbach, M. Stark, and B. Schiele. Evaluating Knowledge Transfer and Zero-Shot Learning in a Large-Scale Setting. CVPR, 2011.

- アトリビュートと物体，物体間の関係性をwebから発見
  - 811アトリビュート：WordNetから
  - Wikipedia, Yahoo Holynyms, Yahoo Image, Yahoo Snippers (Yahoo Web)で関係性をマイニング

Approach	Top 5 Error	Top 1 Error
1. One-vs-all		
(=leaf WordNet nodes)	37.6 (2.91)	57.2 (5.77)
2. Hierarchical		
inner WordNet nodes	71.3 (7.31)	90.7 (8.69)
all WordNet nodes	50.4 (5.49)	67.9 (7.54)
leaf nodes, cost sensitive	48.6 (4.71)	60.2 (5.66)
SVM stacking, all nodes	36.8 (2.84)	56.3 (5.59)
3. Attributes		
Wikipedia	63.7 (5.21)	81.5 (8.52)
Yahoo Holonyms	68.7 (5.61)	87.1 (9.24)
Yahoo Image	74.0 (5.80)	90.6 (10.28)
Yahoo Snippets	67.2 (5.33)	84.6 (8.55)
all attributes	56.4 (4.63)	75.9 (7.32)
SVM stacking, all attributes	43.8 (3.38)	63.5 (6.34)

Table 3: Large scale knowledge sharing results. Shown is flat error in % (hierarchical error)

	On 200 unseen classes	
Approach	Top-5 Error	Top-1 Error
1. Hierarchical		
leaf WordNet nodes	72.8 (4.72)	91.3 (11.73)
inner WordNet nodes	66.7 (4.20)	88.7 (11.16)
all WordNet nodes	65.2 (4.10)	88.4 (11.24)
2. Attributes		
Wikipedia	80.9 (5.17)	94.5 (11.69)
Yahoo Holonyms	77.3 (4.91)	94.0 (12.56)
Yahoo Image	81.4 (5.19)	95.5 (12.53)
Yahoo Snippets	76.2 (4.87)	93.3 (11.53)
all attributes	70.3 (4.57)	90.4 (11.62)
3. Direct Similarity		
Wikipedia	75.6 (5.20)	91.8 (11.28)
Yahoo Web	69.3 (4.49)	89.7 (11.10)
Yahoo Image	72.0 (4.60)	90.7 (11.26)
Yahoo Snippets	75.5 (4.89)	91.6 (11.27)
all measures	66.6 (4.41)	88.4 (10.65)

Table 4: Zero-shot recognition. Flat error in % (hierarchical error).

まとめ

- CVPR2011における物体・シーン認識のトレンド（の一部）について紹介した

- データセットバイアス
  - データセットのクオリティ評価
  - ドメイン適応の重要性

- 転移学習
  - ドメイン適応，マルチタスク学習，知識転移

キーワード

- 大規模　Large scale
  - データセットのバイアス
  - 実世界を表現するための多様性
- 疎表現　Sparse representation
  - 石を投げればsparsenessに当たる，，，
- 属性　Attribute
  - 見慣れない物体の記述
  - Zero-shot認識，知識転移
  - 物体識別を補助する中間特徴
- 転移学習　Transfer learning
  - マルチタスク学習，知識転移，ドメイン適合
- 深い構造　Deep architecture
  - Deep learning, feature learning

少なくとも自分の心構え

- 論文1本に対して関連論文は50本以上読んでいるか？
  - 研究の価値＝（コンテンツ）×（表現）
  - ノンネーティブ：データドリブンアプローチ
- 新たな問題を提起しているか？
  - 従来研究から生じる新たな疑問は何か？
  - 疑問点を投げかけることで読者にその問題に集中させる.
- コントリビューションは明確か？
  - Our contributions are …と箇条書きに.
  - 小さな差分でも自信を持って記述.
  - 本当にインクリメンタルでない研究はあるのか？
- 実験は十分か？
  - 定理の証明のような事実の積み重ね以外の研究は実験による検証が大切.
  - 感覚的には論文の半分は実験.
  - 査読の体の良い断り方：なかなかいいけど，実験が足りないよ.