Transfer Learning: Introduction & Application

Yun Gu

Department of Automation
Shanghai Jiao Tong University
Shanghai, CHINA

June 16th, 2014

Outline

Outline of this section

What is Transfer Learning?[Pan and Yang, 2010]

Naive View (Transfer Learning)

Transfer Learning (i.e. Knowledge Transfer, Domain Adaption) aims at applying knowledge learned **previously** *to solve* **new** *problems faster or with better solutions.*

What is Transfer Learning?[Pan and Yang, 2010]

Naive View (Transfer Learning)

Transfer Learning (i.e. Knowledge Transfer, Domain Adaption) aims at applying knowledge learned **previously** *to solve* **new** *problems faster or with better solutions.*

Transfer

- What to "Transfer"?
- How to "Transfer"?
- When to "Transfer"?

Learning

- Machine Learning Scheme.
- Relationship with other ML tech?

What is Transfer Learning?[Pan and Yang, 2010]

Naive View (Transfer Learning)

Transfer Learning (i.e. Knowledge Transfer, Domain Adaption) aims at applying knowledge learned **previously** *to solve* **new** *problems faster or with better solutions.*

Transfer

- What to "Transfer" ?
- How to "Transfer" ?
- When to "Transfer" ?

Learning

- Machine Learning Scheme.
- Relationship with other ML tech?

In Top-Level Conference: 7 papers in CVPR 2014 are related with Transfer Learning;
In Top-Level Journal: M.Guilaumin, et.al, "ImageNet Auto-Annotation with Segmentation Propagation", IJCV,2014

What is Transfer Learning? (Cont.)

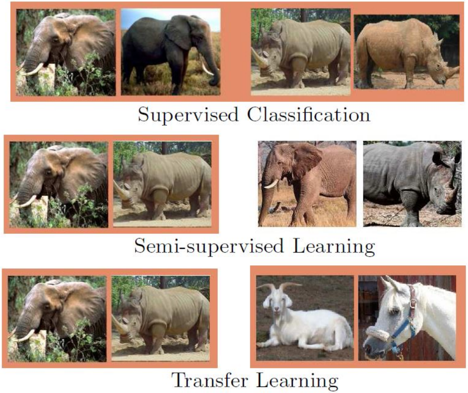


Figure : Supervised classification uses labeled examples of elephants and rhinos; semi-supervised learning uses additional unlabeled examples of elephants and rhinos; transfer learning uses additional labeled datasets[Raina et al., 2007].

Motivation

Why we need Transfer Learning[Tang et al., 2012]?

- **Labeled data are expensive and limited.**
- **Related data are cheap and sufficient.**

Motivation

Why we need Transfer Learning[Tang et al., 2012]?

- **Labeled data are expensive and limited.**
- **Related data are cheap and sufficient.**

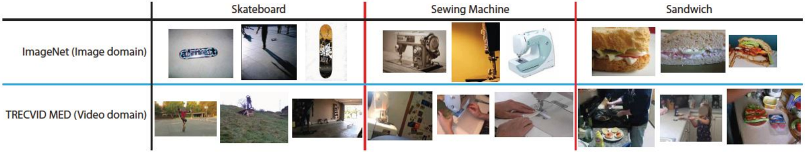


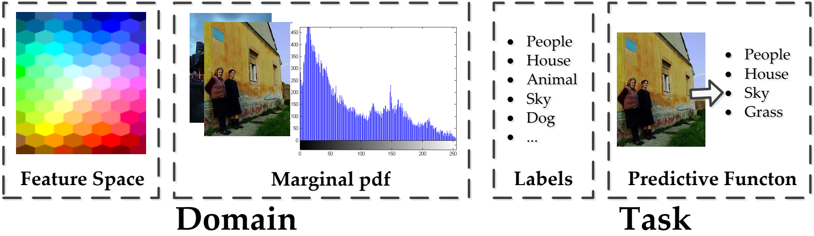
Figure : Object detector for static image is easy to obtain. However, the labeled data for video task are limited and expensive.

Terminologies

- **Domain**: A domain $\mathcal{D} = \{\mathcal{X}, P(X)\}$ consists of two components: a feature space $\mathcal{X}$ and a marginal prob distribution $P(X), X \in \mathcal{X}$.
- **Task**: Given a spefic domain, a task $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$ also consists of two components: a label space $\mathcal{Y}$ and the predictive function $f(\cdot) = P(y|x)$. *The predictive is unknown for us but can be learned from training data, which consists of data pair* (x_i, y_i).

Terminologies

- **Domain**: A domain $\mathcal{D} = \{\mathcal{X}, P(X)\}$ consists of two components: a feature space $\mathcal{X}$ and a marginal prob distribution $P(X), X \in \mathcal{X}$.
- **Task**: Given a spefic domain, a task $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$ also consists of two components: a label space $\mathcal{Y}$ and the predictive function $f(\cdot) = P(y|x)$. *The predictive is unknown for us but can be learned from training data, which consists of data pair* (x_i, y_i).



Terminologies (Cont.)

- **Source/Target Domain Data**: a set of labeled data D_S and D_T

  $$D_S = \{(x_{S_1}, y_{S_1})...(x_{S_{n_S}}, y_{S_{n_S}})\}, D_T = \{(x_{T_1}, y_{T_1})...(x_{T_{n_T}}, y_{T_{n_T}})\}$$

  In most cases, $0 \leq n_T \ll n_S$.

Terminologies (Cont.)

- **Source/Target Domain Data**: a set of labeled data D_S and D_T

  $$D_S = \{(x_{S_1}, y_{S_1})...(x_{S_{n_S}}, y_{S_{n_S}})\}, D_T = \{(x_{T_1}, y_{T_1})...(x_{T_{n_T}}, y_{T_{n_T}})\}$$

  In most cases, $0 \leq n_T \ll n_S$.

Definition (Transfer Learning)

Given a source domain $\mathcal{D}_S$ and learning task $\mathcal{T}_S$, a target domain $\mathcal{D}_T$ and learning task $\mathcal{T}_T$, **transfer learning** *aims to help improve the learning of the target predictive function $f_T(\cdot)$ in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$.*

Outline of this section

1. Overview of Transfer Learning
2. Categorization
   - Three Research Issues
   - Different Settings
3. Applications
4. Conclusion

Three Research Issues

- **"What to transfer?"**
  Which part of knowledge can be transferred across domain?
  e.g. feature representation, parameter settings, latent feature distribution, etc.
- **"How to transfer?"** Specific learning algorithms to transfer the knowledge.
  e.g. TrAdaBoost (*Dai,2007*), Structural Correspondence Learning (*Blitzer,2006*),etc.
- **"When to transfer?"** Asks in which situations, transferring skills should be done. Likewise, we are interested in knowing in which situations, knowledge should not be transferred.

Categorization of Transfer Learning

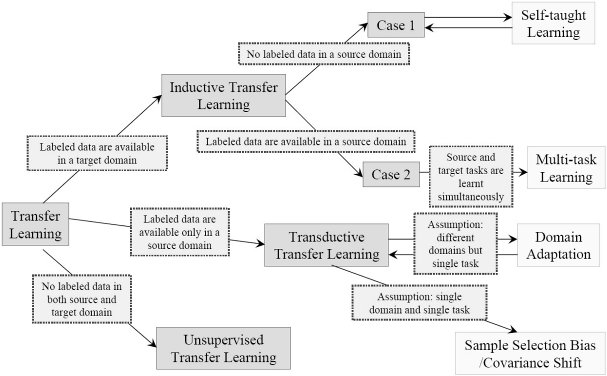


Figure : Different settings of Transfer Learning based on the availability of Source/Target Domain Labels.

Outline of this section

1. Overview of Transfer Learning
2. Categorization
3. Applications
   - Image Annotation
   - Image Classification
   - Deep Learning
4. Conclusion

Transfer Learning: Image Annotation

Yun Gu

Knowledge Transfer for Automatic Image Annotation

Main Idea: Learning a new class(target) is helped by labeled examples of other related classes (source). This is actually a Inductive Transfer Learning scheme.

Knowledge Transfer for Automatic Image Annotation

Main Idea: Learning a new class(target) is helped by labeled examples of other related classes (source). This is actually a Inductive Transfer Learning scheme.

- **Parameters Transfer**: Use the parameters from the source classifier as a prior for target model.
  e.g. Transfer w in SVM Classifier from source to target. *Aytar, ICCV 2011.*
- **Feature Transfer**: Transfer knowledge through an intermediate attributive layer shared by many classes.
  e.g. The color or basic texture. *Lampert, CVPR 2009.*
- **Transfer between Classes**: Transfer object parts between classes, such as wheels between "car" and "bike". e.g. *Ott, CVPR 2011.*
- **From Annotated to Bounded**: Transfer from the images only annotated by tags to the localization task. e.g. *Guillaumin, CVPR 2012.*

Auto-Annotation with Segmentation Propagation [Guillaumin et al., 2014]

- **Task**: Weakly-supervised segmentation on ImageNet dataset (500k images,577 classes).
- **A new transfer scheme (Window-based Transfer)**:
  - Segmented Images from PASCAL VOC 2010 (1928 images, 4203 objects).
  - Images with bounding boxes from ImageNet (60k images).
  - Images only with tags from ImageNet (440k images).
  - Images to be segmented.

Window Transfer

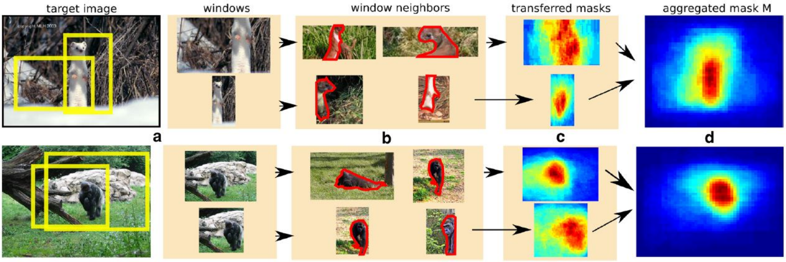


Figure : Examples of window-level segmentation transfer: From segmented to bounded.

Segmentation Propagation

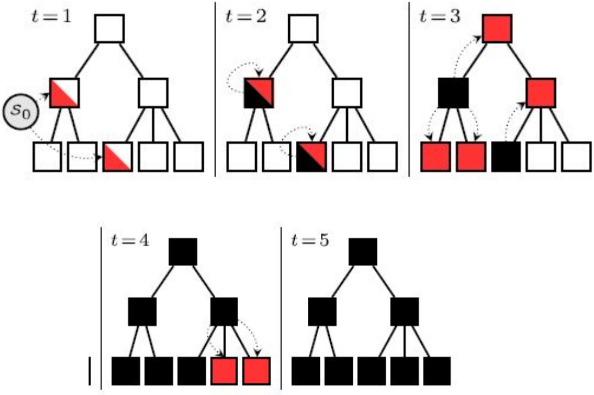
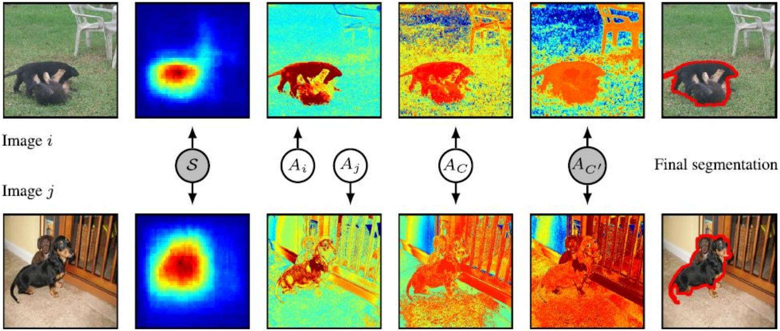


Figure : *white*:"unsegmented";*red*:"being segmented";*black*:"already segmented"

Class-wise Cosegmentation

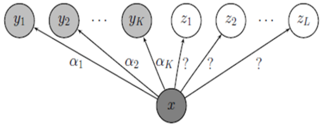


Transfer Learning: Image Classification

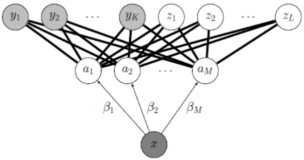
Shaoyong Jia

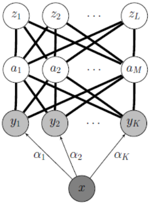
TL for Image Classification [Lampert et al., 2009]

- **Problem**: Object classification when training and test classes are disjoint.



- **Proposal**: Transfer learning for object detection by between class attributes.

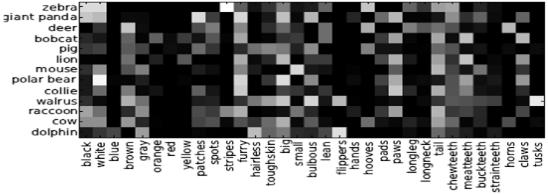




Transfer Learning for Image Classification (Cont.)

- **Attribute**: Human-specified high-level description, which consists of arbitrary semantic attributes, like shape, color or even geographic information.

- **The class-attribute matrices**



Transfer Learning for Image Classification (Cont.)

- **Dataset**: Animals with Attributes of over 30475 animal images, 85 semantic attributes and 50 classes with at minimum of 92 images for any class[1].

- Base package (1M) including the class/attribute table:AwA-base.tar.bz2 (everybody needs this)
- Color Histogram features (124M): AwA-features-cq.tar.bz2
- Local Self-Similarity features (30M): AwA-features-lss.tar.bz2
- PyramidHOG (PHOG) features (28M): AwA-features-phog.tar.bz2
- SIFT features (44M): AwA-features-sift.tar.bz2
- colorSIFT features (44M): AwA-features-rgsift.tar.bz2
- SURF features (49M): AwA-features-surf.tar.bz2
- DECAF features (122M): AwA-features-decaf.tar.bz2 **(NEW!)**
- Source code (30K) illustrating DAP and IAP methods: AwA-code.tar.bz2

Addendum: new attributes.py script that work with recent versions of Shogun

- Example Images (15M): AwA-examples.tar.bz2 *(3 example per class, e.g. for illustrative use in publications)*
- Full Image Set in JPEG format: *not directly downloadable for copyright reasons*

please ask at <*chl(at)ist.ac.at*>.

[1] Website Link: `http://attributes.kyb.tuebingen.mpg.de/`

Transfer Learning for Image Classification (Cont.)

- **Implementation**:Use a probabilistic model to reflect the graphical.

DAP
- Image-attribute stage:

$$p(\alpha|x) = \prod_{m=1}^{M} p(\alpha_m|x)$$

- Attribute-class stage:

$$p(z|\alpha) = \frac{p(z)}{p(\alpha^z)}[\alpha = \alpha^z]$$

- Image-class stage:

$$p(z|x) = \sum_{\alpha \in \{0,1\}^M} p(z|\alpha)p(\alpha|x) = \frac{p(z)}{p(\alpha^z)} \Pi_{m=1}^{M} p(\alpha_m^z|x)$$

- Decision rule:

$$f(x) = argmax_{l=1,\dots,L} \prod_{m=1}^{M} \frac{p(\alpha_m^{z_l}|x)}{p(\alpha_m^{z_l})}$$

IAP
- Image-attribute stage:$p(\alpha_m|x) = \sum_{k=1}^{K} p(\alpha_m|y_k)p(y_k|x)$
- Other stages are in the same way in DAP:

Transfer Learning for Image Classification (Cont.)

- Experimental design:6180 images of 10 classes for test while 24295 images of 40 classes for training.
- Results:Accuracy of 40.5% for DAP while 27.8% for IAP .

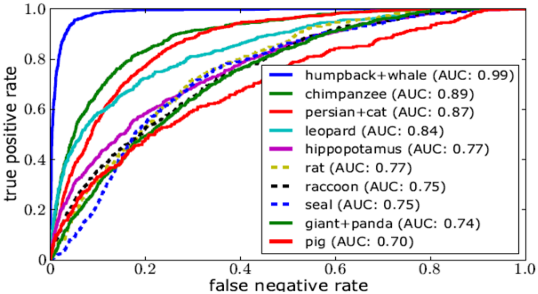


Figure : Note: Detection performance of object classification with disjoint training and test classes(DAP method):ROC-curves and area under curve(AUC) for the 10 Animals with Attributes test classes.

Transfer Learning for Image Classification (Cont.)

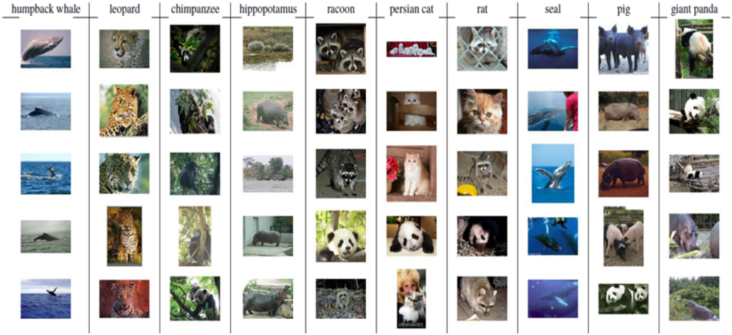


Figure : Note: The five images with highest posterior score for each test class.

Transfer Learning: Combined with Deep Learning

Haoyang Xue

The Characteristics of Deep Learning

Advantages
- Outstanding classification performance in large-scale visual recognition challenge

Flaws
- Numerous parameters;
- A large scale number of annotated samples needed;
- Time consuming.

The Characteristics of Deep Learning

Advantages
- Outstanding classification performance in large-scale visual recognition challenge

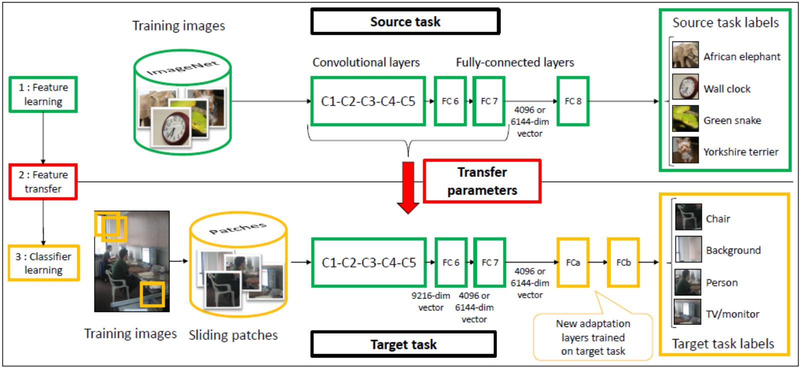
Flaws
- Numerous parameters;
- A large scale number of annotated samples needed;
- Time consuming.

Solution
- The multilayer networks of a training set include many intermediate features or presentations.(Avoid training large part of the network for a new task)
- Can we just transfer the middle presentations of a pre-trained network on one dataset to new dataset for new target task?(Avoid collecting a large scale of training data)

Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks[Oquab et al., 2013]

- In this paper, the author propose an approach to re-use a pre-trained CNN network to a new dataset and estimate its performance in target tasks.
- A pre-trained CNN network on the source dataset(ImageNet) for classification task.
- A adaptation layer is then trained with the data in the new dataset(Pascal VOC)to solve the differences between two tasks.
- The new network is applied to the object classification task on VOC2007 and VOC2012 test set.

The Main Framework



The Performance

	plane	bike	bird	boat	btl	bus	car	cat	chair	cow	table	dog	horse	moto	pers	plant	sheep	sofa	train	tv	mAP
INRIA [32]	77.5	63.6	56.1	71.9	33.1	60.6	78.0	58.8	53.5	42.6	54.9	45.8	77.5	64.0	85.9	36.3	44.7	50.6	79.2	53.2	59.4
NUS-PSL [44]	82.5	79.6	64.8	73.4	**54.2**	75.0	77.5	79.2	46.2	62.7	41.4	74.6	**85.0**	76.8	91.1	53.9	61.0	**67.5**	83.6	70.6	70.5
PRE-1000C	**88.5**	**81.5**	**87.9**	**82.0**	47.5	**75.5**	**90.1**	**87.2**	**61.6**	**75.7**	**67.3**	**85.5**	83.5	**80.0**	**95.6**	**60.8**	**76.8**	58.0	**90.4**	**77.9**	**77.7**

Table 1: Per-class results for object classification on the VOC2007 test set (average precision %).

	plane	bike	bird	boat	btl	bus	car	cat	chair	cow	table	dog	horse	moto	pers	plant	sheep	sofa	train	tv	mAP
NUS-PSL [49]	**97.3**	**84.2**	80.8	**85.3**	**60.8**	**89.9**	**86.8**	89.3	**75.4**	77.8	**75.1**	83.0	87.5	**90.1**	95.0	57.8	79.2	**73.4**	**94.5**	**80.7**	82.2
NO PRETRAIN	85.2	75.0	69.4	66.2	48.8	82.1	79.5	79.8	62.4	61.9	49.8	75.9	71.4	82.7	93.1	59.1	69.7	49.3	80.0	76.7	70.9
PRE-1000C	93.5	78.4	*87.7*	80.9	57.3	85.0	81.6	*89.4*	66.9	73.8	62.0	*89.5*	83.2	87.6	*95.8*	*61.4*	79.0	54.3	88.0	78.3	78.7
PRE-1000R	93.2	77.9	83.8	80.0	55.8	82.7	79.0	84.3	66.2	71.7	59.5	83.4	81.4	84.8	95.2	59.8	74.9	52.9	83.8	75.7	76.3
PRE-1512	94.6	82.9	**88.2**	84.1	60.3	89.0	84.4	**90.7**	72.1	**86.8**	69.0	**92.1**	**93.4**	88.6	**96.1**	**64.3**	**86.6**	62.3	91.1	79.8	**82.8**

Table 2: Per-class results for object classification on the VOC2012 test set (average precision %).

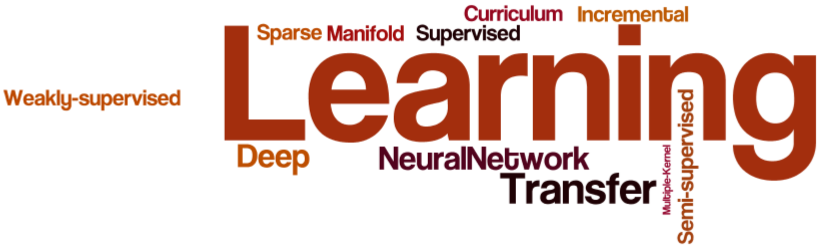
- A simple transfer learning procedure yields state-of-the-art results on challenging benchmark datasets of much smaller size.

Outline of this section

Conclusion

Transfer Learning is still a hot topic in various fields.

- Image Annotation/Classification;
- Text Classification;
- Recommendation;
- Software Engineering;
- etc.



References I

Guillaumin, M., K U Ttel, D., and Ferrari, V. (2014).
Imagenet auto-annotation with segmentation propagation.
International Journal of Computer Vision, pages 1–21.

Lampert, C. H., Nickisch, H., and Harmeling, S. (2009).
Learning to detect unseen object classes by between-class attribute transfer.
In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 951–958. IEEE.

Oquab, M., Bottou, L., Laptev, I., Sivic, J., et al. (2013).
Learning and transferring mid-level image representations using convolutional neural networks.

Pan, S. J. and Yang, Q. (2010).
A survey on transfer learning.
Knowledge and Data Engineering, IEEE Transactions on, 22(10):1345–1359.

Raina, R., Battle, A., Lee, H., Packer, B., and Ng, A. Y. (2007).
Self-taught learning: transfer learning from unlabeled data.
In *Proceedings of the 24th international conference on Machine learning*, pages 759–766. ACM.

Tang, K., Ramanathan, V., Li, F.-F., and Koller, D. (2012).
Shifting weights: Adapting object detectors from image to video.
volume 1, page 2.

Thank you.
Q&A.