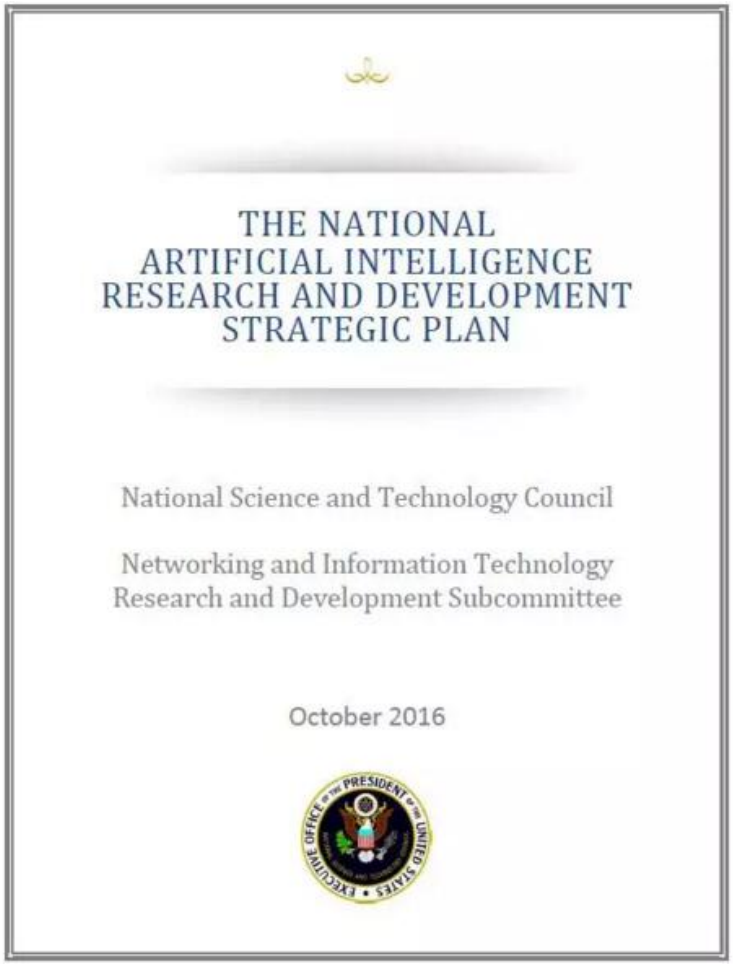
Human-like Visual Learning

Peng Cui, Wenwu Zhu

Tsinghua University

National Strategies for AI

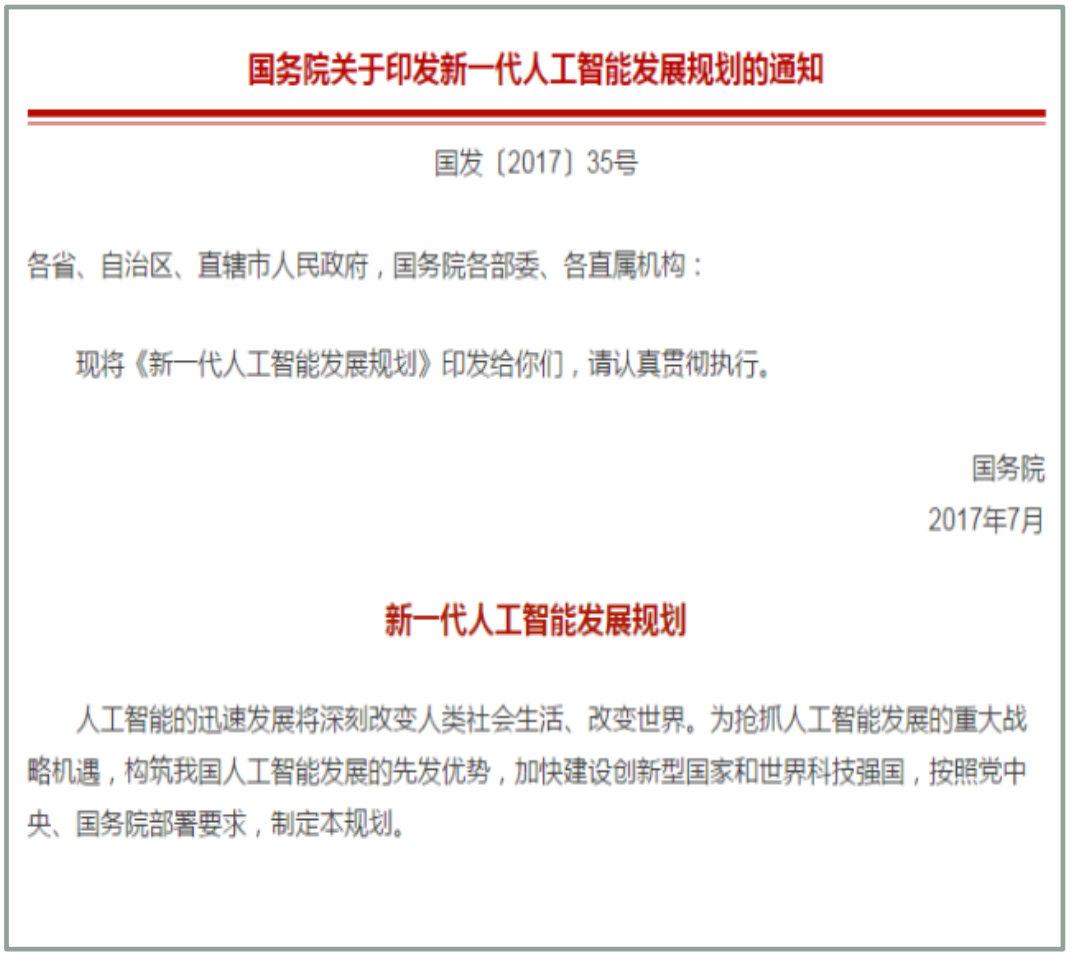


THE NATIONAL
ARTIFICIAL INTELLIGENCE
RESEARCH AND DEVELOPMENT
STRATEGIC PLAN

National Science and Technology Council

Networking and Information Technology
Research and Development Subcommittee

October 2016



国务院关于印发新一代人工智能发展规划的通知

国发〔2017〕35号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

现将《新一代人工智能发展规划》印发给你们，请认真贯彻执行。

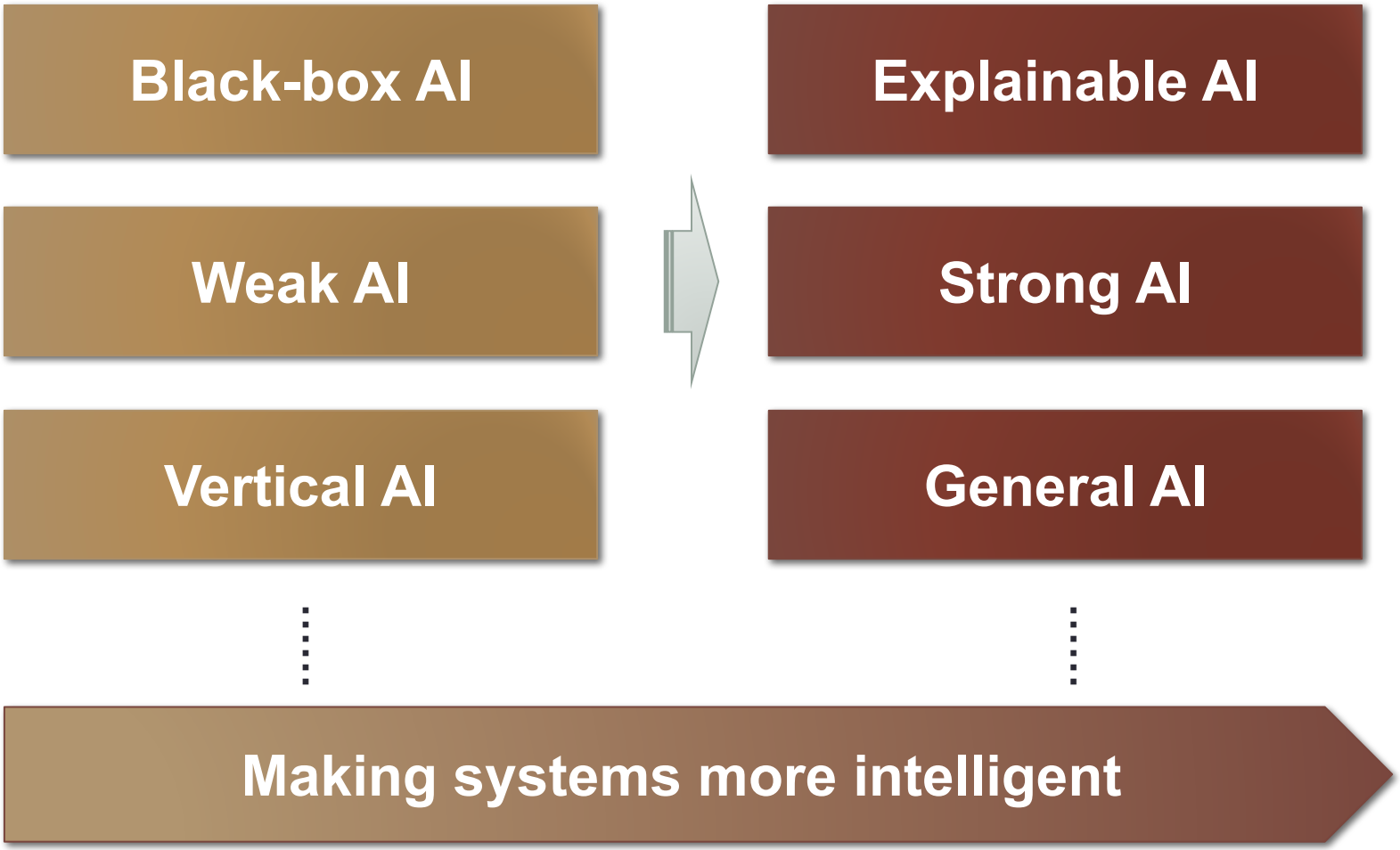
国务院

2017年7月

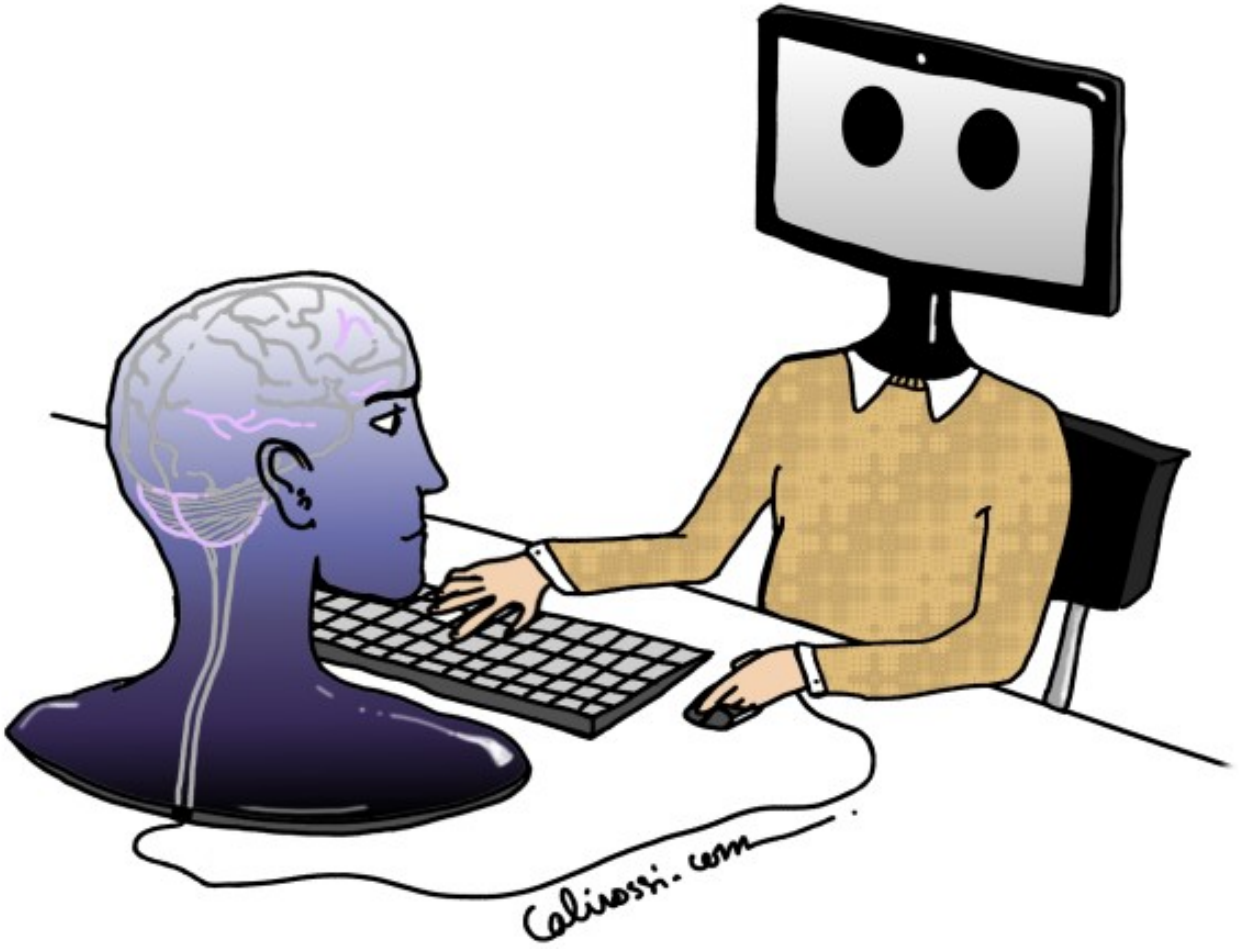
新一代人工智能发展规划

人工智能的迅速发展将深刻改变人类社会生活、改变世界。为抢抓人工智能发展的重大战略机遇，构筑我国人工智能发展的先发优势，加快建设创新型国家和世界科技强国，按照党中央、国务院部署要求，制定本规划。

Towards AI 2.0



Reference Model: Human



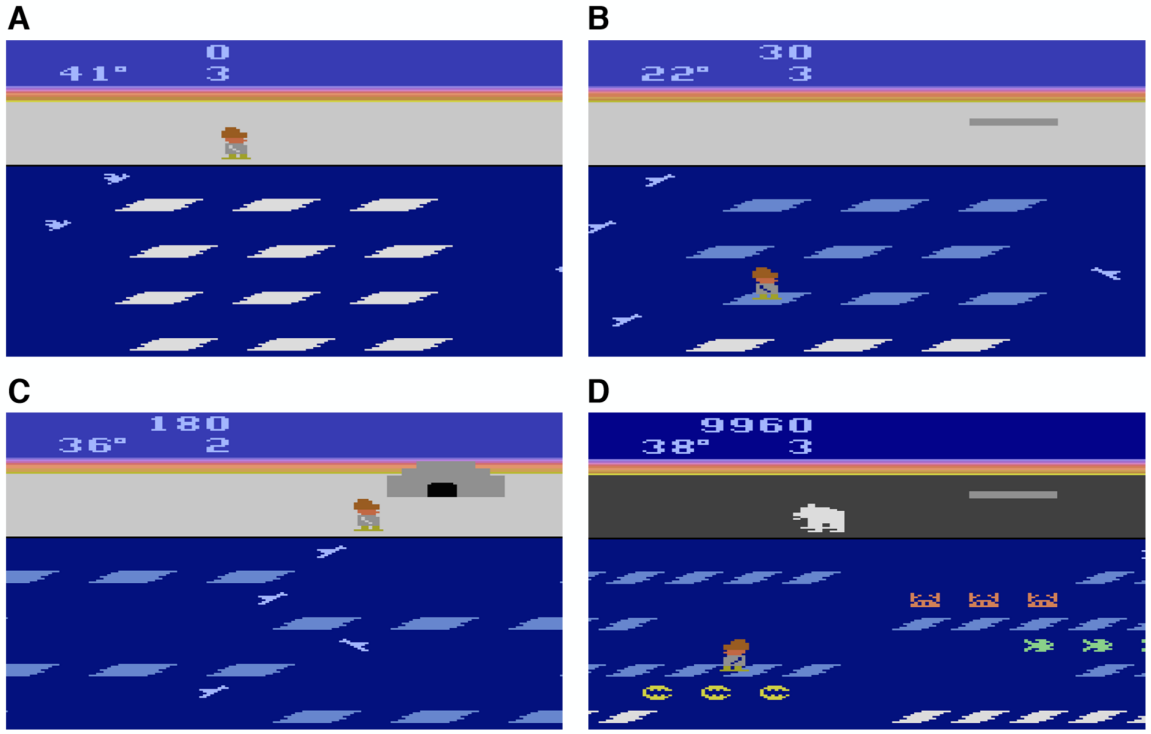
Machine Learning v.s. Human Learning

- Learning Simple Visual Concepts
  - People learn from fewer examples
  - People learn richer representations
  - People can learn to recognize a new character from a single example
  - People learn a concept – a model of the class that allows their acquired knowledge to be flexible applied in new ways.

Lake, Brenden M., et al. “Building machines that learn and think like people.” *Behavioral and Brain Sciences* (2016): 1-101.

Machine Learning v.s. Human Learning

- The Frostbite Challenge



Lake, Brenden M., et al. “Building machines that learn and think like people.” *Behavioral and Brain Sciences* (2016): 1-101.

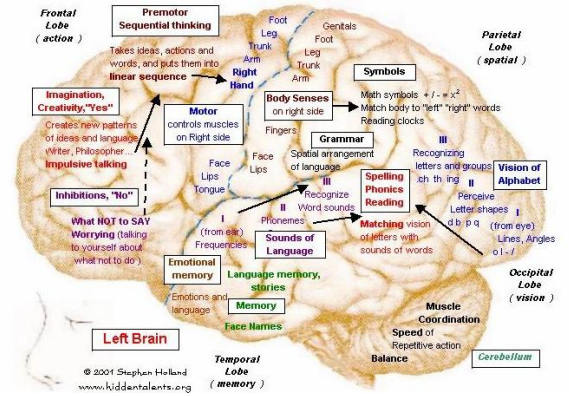
Machine Learning v.s. Human Learning

- The Frostbite Challenge
  - Optimal Solution up till now: Deep Q Network
  - Shortcomings compared with humans
    - People use less time to practice to reach nearly the same average score: human for 2 hours and DQN for 924 hours.
    - Human could grasp the basics of the game just after a few minutes of playing.
    - If humans are able to watch an expert playing for a few minutes, they can learn even faster.
    - Humans are more flexible, i.e. after they learn how to play, they could finish arbitrary new tasks and goals. (e.g. get closest to score 300 etc.)

Lake, Brenden M., et al. “Building machines that learn and think like people.” *Behavioral and Brain Sciences* (2016): 1-101.

Technical Paths of Learning from Human

Brain-like Learning



Bottom-Up

New AI algorithms

Depends on how much we can understand human brain.
The computer architecture v.s. brain architecture?

Technical Paths of Learning from Human

Human-like Learning

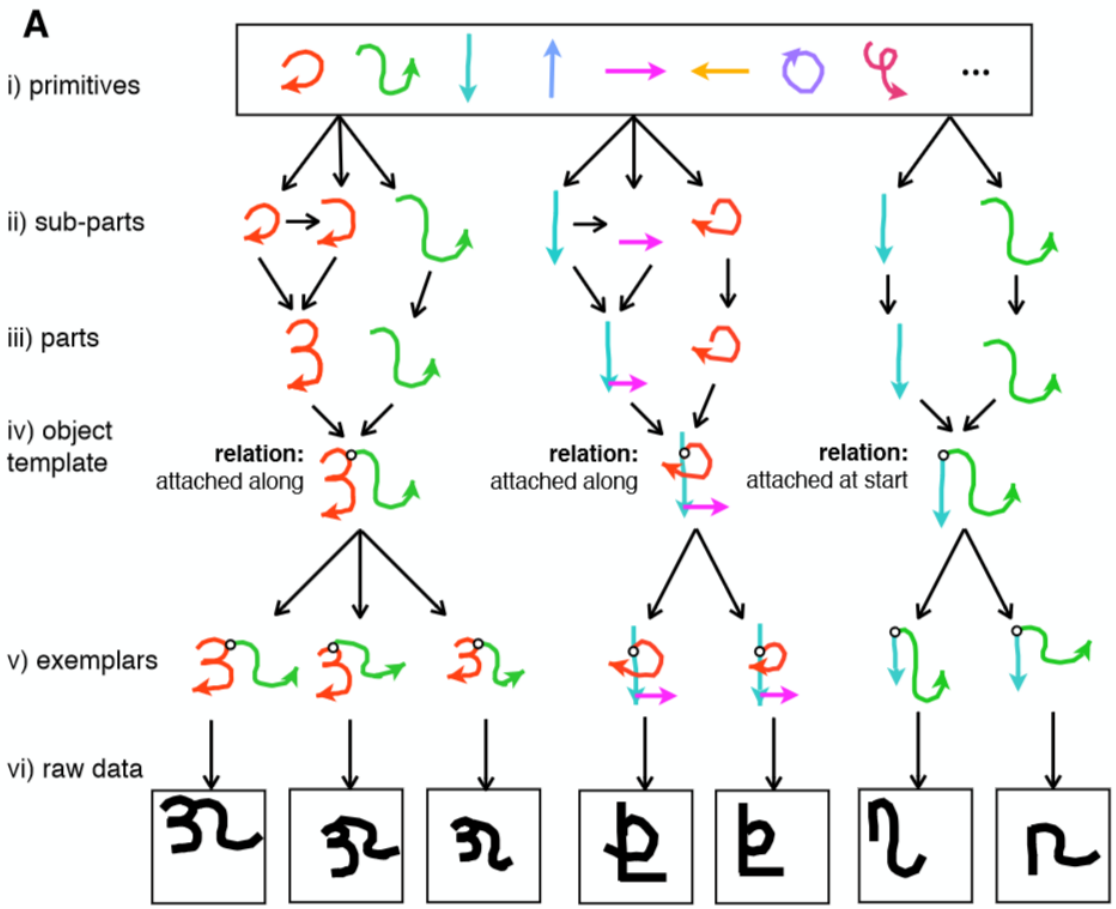
Inference, association, imagine …

Top-down

New AI algorithms

Another path for inventing new learning mechanisms.

A Successful Example



Inference: compositionality and causality

Association: Learning to learn

Far less samples are required for training than deep models.

Simple showcase

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. Science, 350 (6266), 1332-1338.

Human-like Visual Learning

From human:
Can we learn from how people learn visual objects?

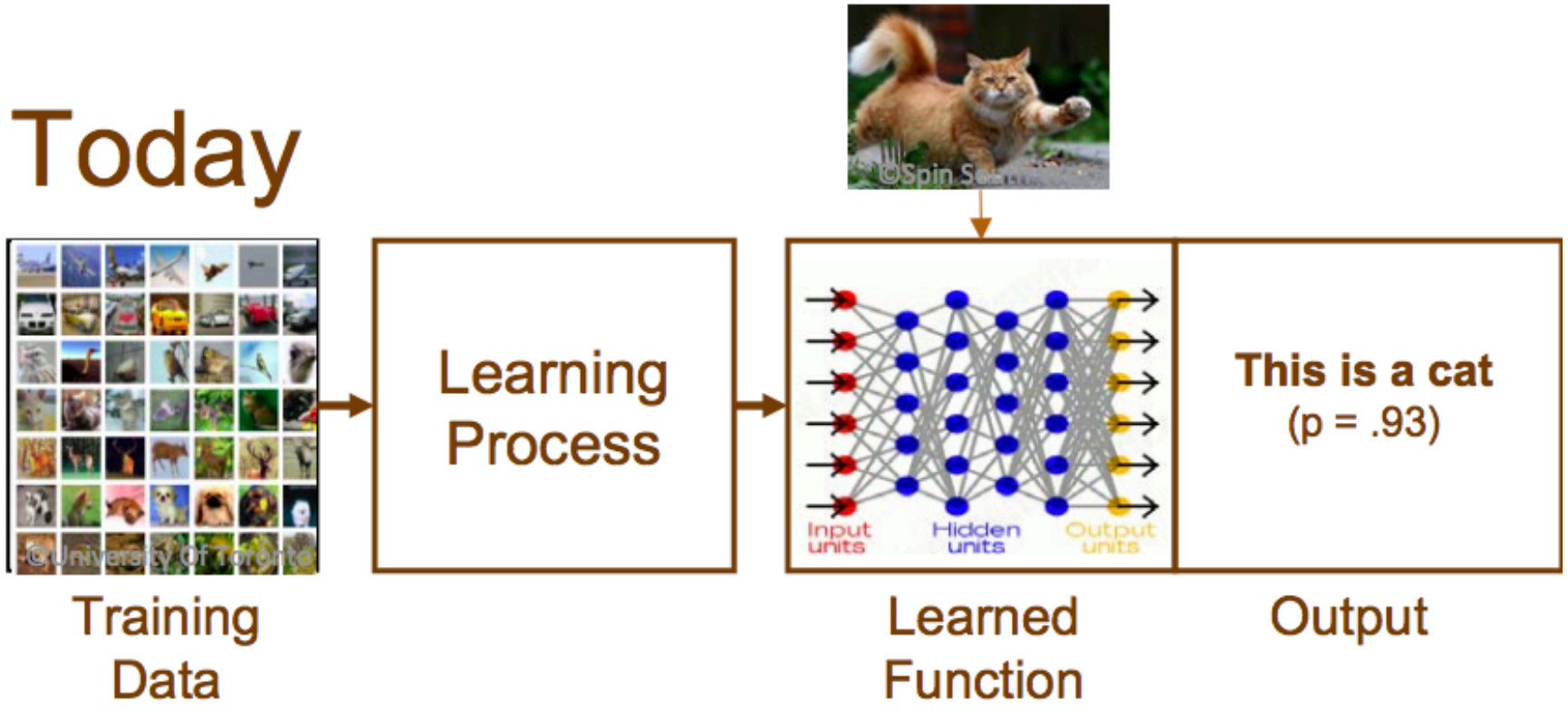
For human:
Can we infer how people behave with visual objects?

Human-like Visual Learning

From human:

Can we learn from how people learn visual objects?

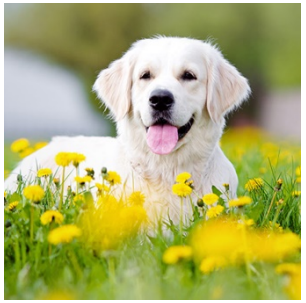
Deep Learning has dominated the visual world

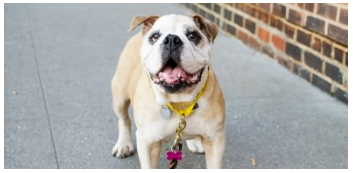


Deep learning has indeed advanced the visual learning significantly.

What are the downsides?

(1) Simple Hypothesis: I.I.D.





FAIL!

Worse case: small training samples?

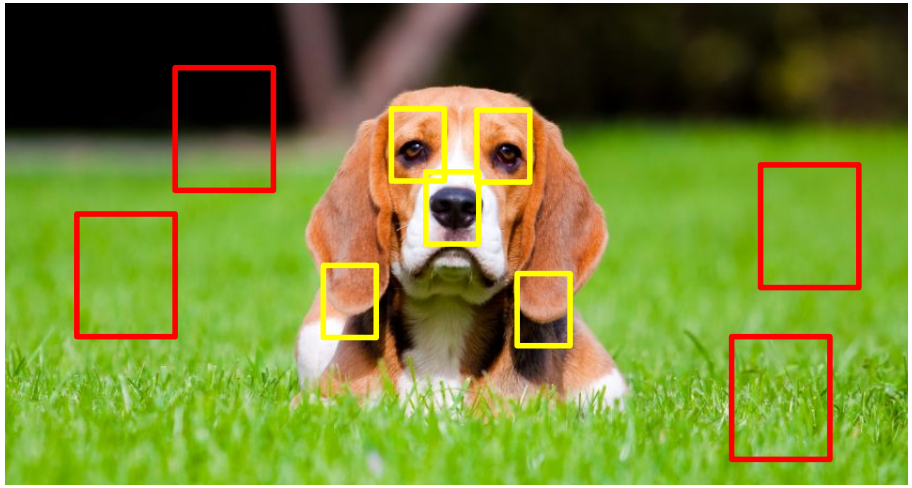
What are the downsides?

(1) Simple Hypothesis: I.I.D.

-Do we **human** have the same problem?

-NO! We have strong **inference** ability.

Correlation features v.s. Causal features



Towards image classification: Correlation v.s. Causality?

What are the downsides?

(2) Inept learning way

dog

wolf

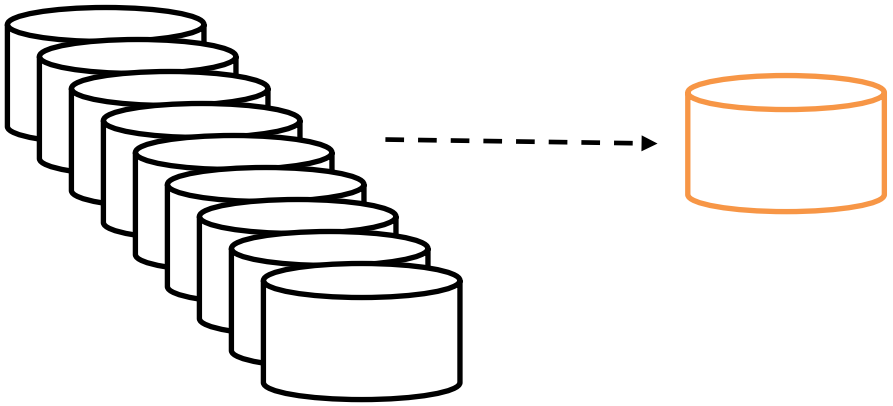
What is the correct way of learning to recognize wolf, given that you have been able to recognize dog?

What are the downsides?

(2) Inept learning way

-How do we **Human** learn a new concept?

-We learn by **association**.



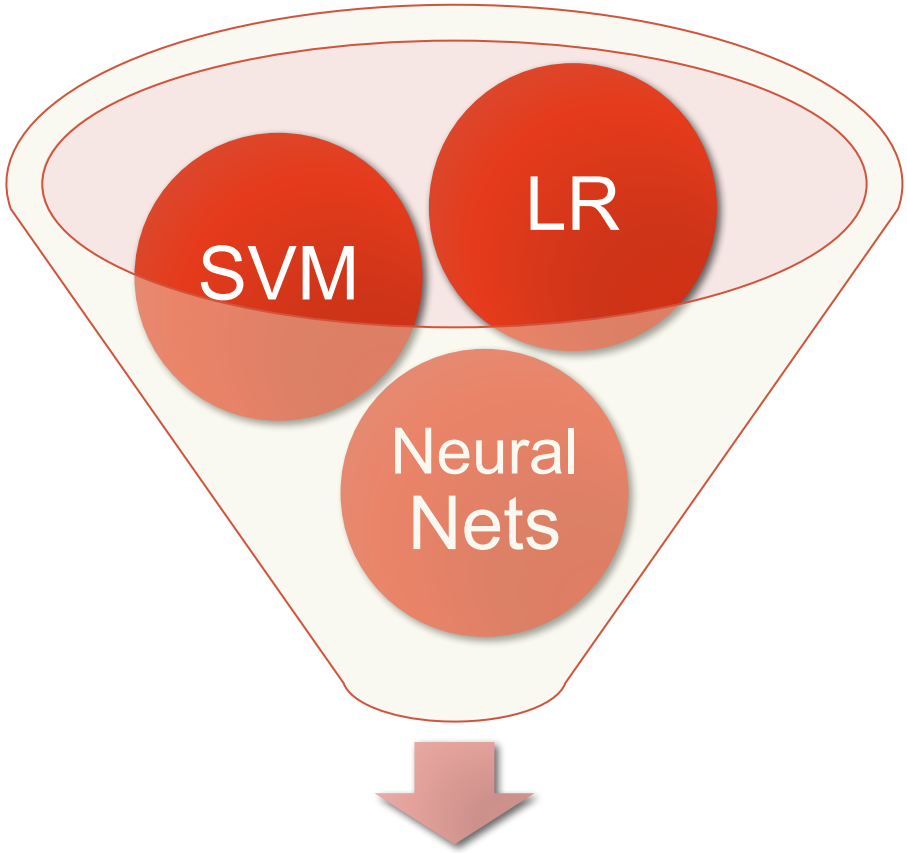
Base Classes Target Class

The more we have learned, the faster we should be able to learn new things.

Learning to learn new concepts

Towards image classification: Correlation v.s. Causality?

Correlation



Correlation Models

These methods excel at leveraging the statistical dependence (*correlation*) between pixels and image label through training data

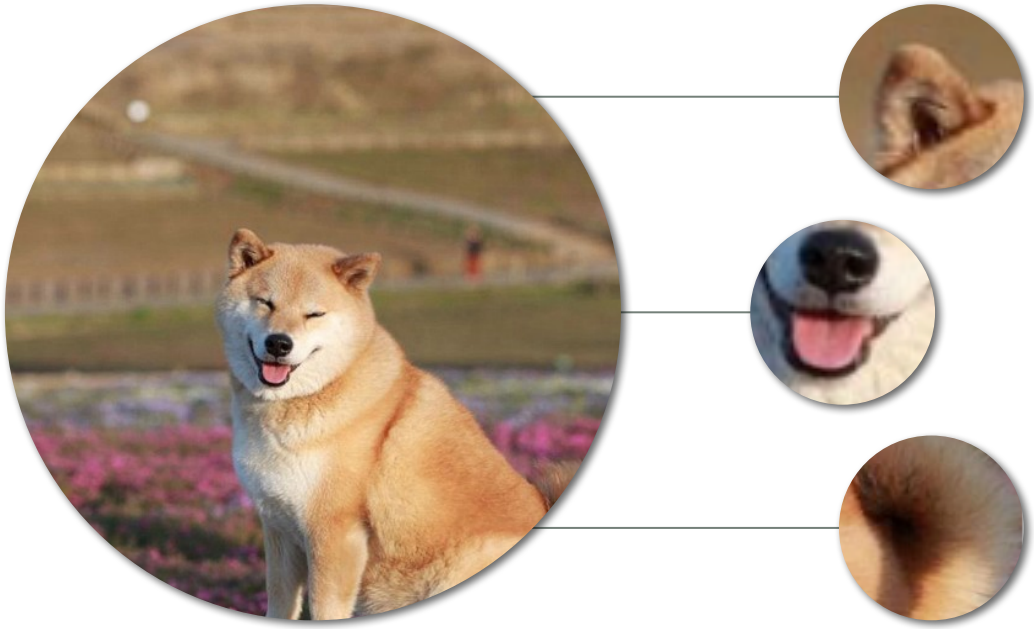
Correlation may not still be held when distribution changes (*non-i.i.d*)

Causality

Causal Inference

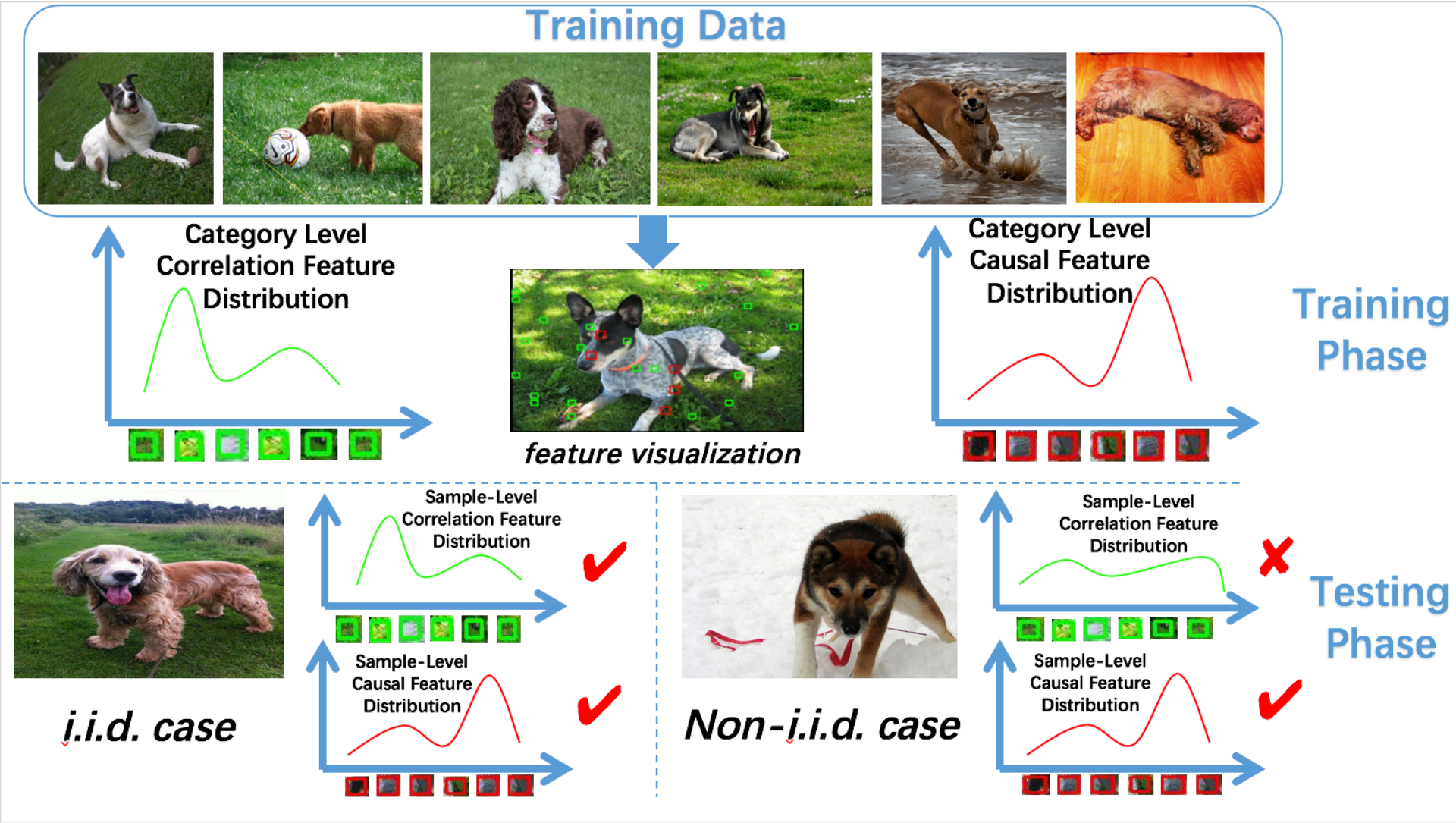
Why is there a dog in this image?
What causes the object to be a dog?

We want to identify those causal features that are stable among different distributions

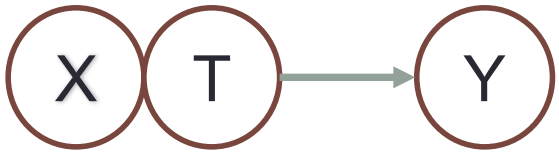


Causal Features

Correlation v.s. Causality

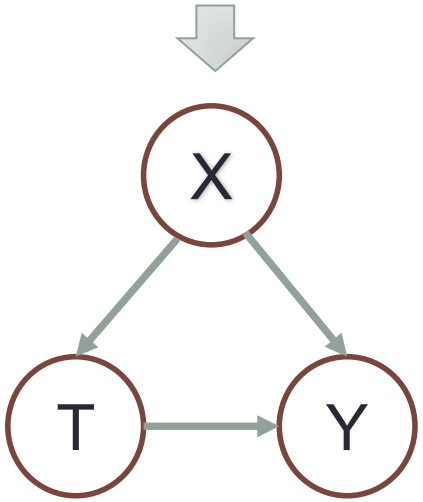


Causal Inference



Typical Correlation Framework

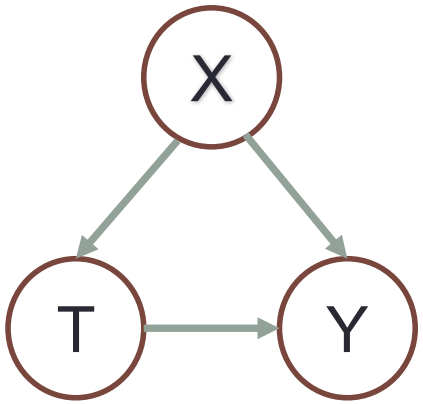
Estimate the correlation effect of variable *T* and output *Y* without evaluating the relationships between *X* and *T*.



Typical Causal Framework

Estimate the causal effect of treatment *T* on output *Y* under the confounder *X* (A/B Testing)

Causal Inference by Absolute Matching



Typical Causal Framework

Analogy of A/B Testing

Given a visual feature T (e.g. a visual word)

↓

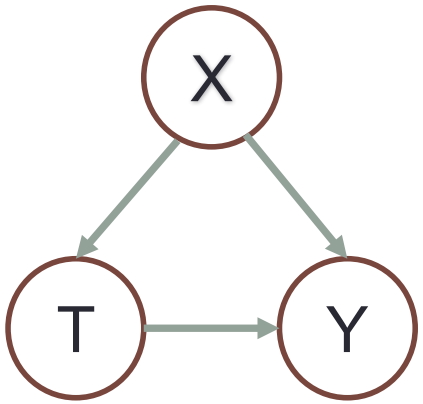
Find out the image pairs that one contains T while the other don't, but they are similar in all other visual features.

↓

Calculate the difference of Y distribution in treated and controlled groups. (correlation between T and Y)

The requirement is too strong and we can hardly find satisfied image pairs.

Causal Inference by Confounder Balancing



Typical Causal Framework

Analogy of A/B Testing

Given a visual feature T (e.g. a visual word)

↓

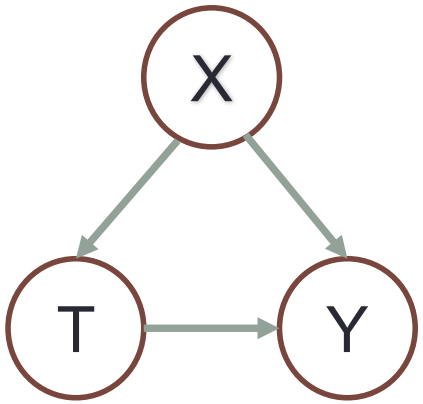
Assign different weights to samples so that the samples with T and the samples without T have similar distributions in X

↓

Calculate the difference of Y distribution in treated and controlled groups. (correlation between T and Y)

Too many parameters. For N samples and K feature, we need to learn K*N weights.

Causal Inference by Global Balancing



Typical Causal Framework

Analogy of A/B Testing

Given **ANY** visual feature T (e.g. a visual word)

↓

Assign different weights to samples so that the samples with T and the samples without T have similar distributions in X

↓

Calculate the difference of Y distribution in treated and controlled groups. (correlation between T and Y)

Reduce the parameter number from K*N to N.

Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

Causal Regularizer

Set image feature *j* as treatment variable

$$\sum_{j=1}^{p} \left\| \frac{X_{-j}^T \cdot (W \odot I_j)}{W^T \cdot I_j} - \frac{X_{-j}^T \cdot (W \odot (1 - I_j))}{W^T \cdot (1 - I_j)} \right\|_2^2,$$

All features excluding treatment *j*

Sample Weights

Indicator of treatment status

Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

Causally Regularized Logistic Regression

$$\min \quad \sum_{i=1}^{n} W_i \cdot \log(1+\exp((1-2Y_i)\cdot(x_i\beta))),$$

$$s.t. \quad \sum_{j=1}^{p} \left\| \frac{X_{-j}^T \cdot (W \odot I_j)}{W^T \cdot I_j} - \frac{X_{-j}^T \cdot (W \odot (1-I_j))}{W^T \cdot (1-I_j)} \right\|_2^2 \leq \lambda_1,$$

$$W \geq 0, \quad \|W\|_2^2 \leq \lambda_2, \quad \|\beta\|_2^2 \leq \lambda_3, \quad \|\beta\|_1 \leq \lambda_4,$$

$$(\sum_{k=1}^{n} W_k - 1)^2 \leq \lambda_5,$$

Sample reweighted logistic loss

Causal Contribution

Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

Dataset

- Source: ***YFCC100M***
- Type: high-resolution and multi-tags
- Scale: 10-category, each with nearly 1000 images
- Method: select 5 ***context tags*** which are frequently co-occurred with the ***major tag*** (category label)

Experimental Setting

- Radical Context Bias
  - Training and testing set are formed by different contexts
- Moderate Context Bias
  - Training and testing set are formed by same contexts but with different percentages
- Label Composition Bias
  - Percentage of positive and negative samples are different in training and testing set

Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

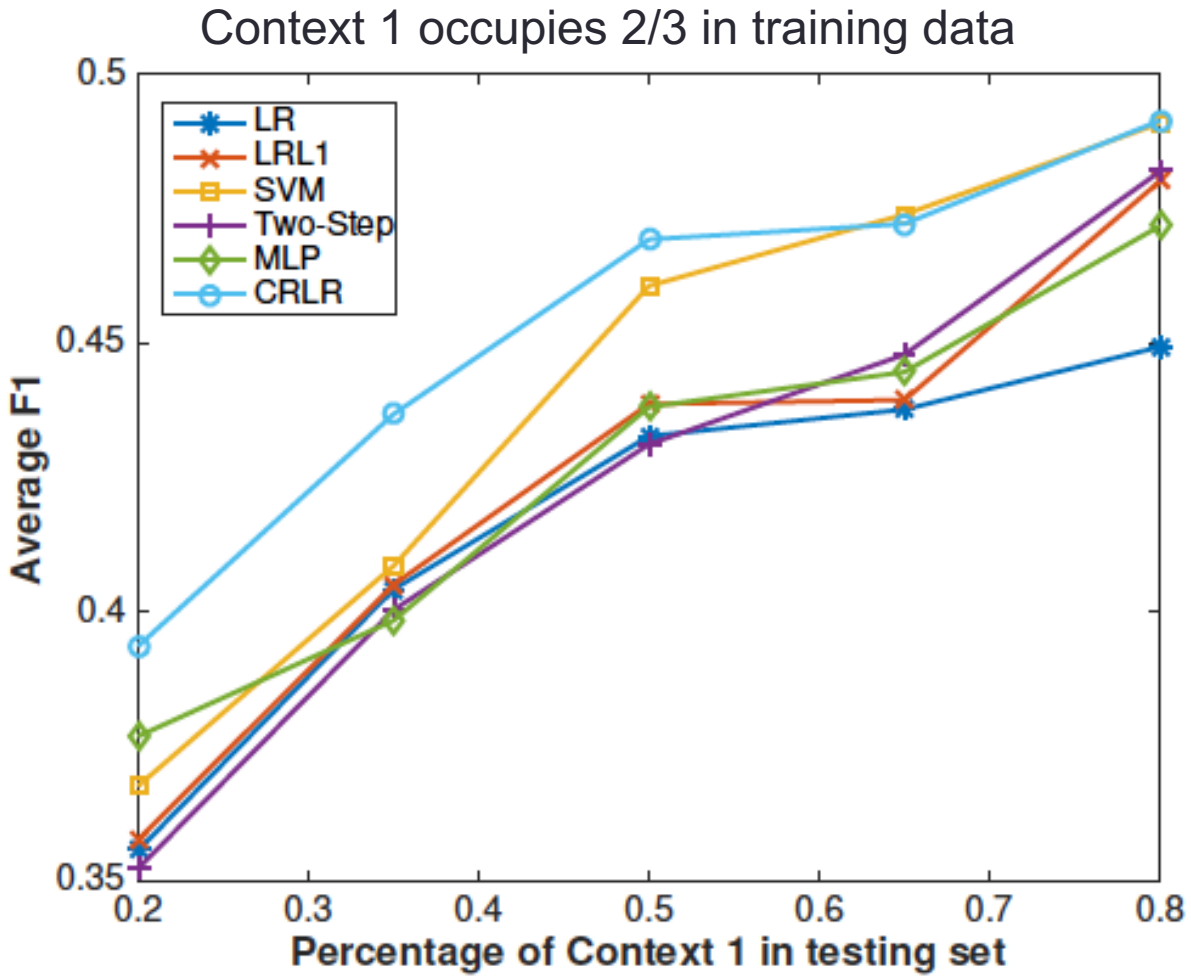
Experimental Result – radical bias

Table 3: Results of classifiers under non-i.i.d. situation with radical context bias in data.

	bird		bridge		car		cat		church	
	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
LR	0.629	0.414	**0.644**	0.450	0.709	0.588	0.617	0.456	0.760	0.637
LR+L_1	0.582	0.283	0.630	0.413	0.692	0.559	0.609	0.424	0.699	0.571
SVM	0.612	0.375	0.638	0.446	0.681	0.548	0.615	0.451	0.764	**0.660**
Two-Step	0.584	0.301	0.639	0.405	0.694	0.539	0.605	0.434	0.767	0.512
MLP	0.568	0.379	0.617	0.337	0.708	0.583	0.586	0.523	0.667	0.634
CRLR	**0.657**	**0.564**	0.617	**0.472**	**0.729**	**0.678**	**0.669**	**0.597**	**0.779**	0.633
	dog		flower		horse		train		tree	
	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
LR	0.565	0.370	0.734	0.635	0.580	0.362	0.592	0.398	0.732	0.618
LR+L_1	0.576	0.307	0.718	0.613	0.580	0.321	0.589	0.384	0.697	0.569
SVM	0.586	0.360	0.720	0.629	0.612	0.404	0.624	0.448	0.681	0.550
Two-Step	0.574	0.389	0.724	0.602	0.606	0.238	0.621	0.321	0.693	0.498
MLP	0.579	0.360	0.726	0.611	0.606	0.388	0.617	0.432	0.710	0.573
CRLR	**0.727**	**0.574**	**0.762**	**0.681**	**0.649**	**0.435**	**0.647**	**0.479**	**0.738**	**0.620**

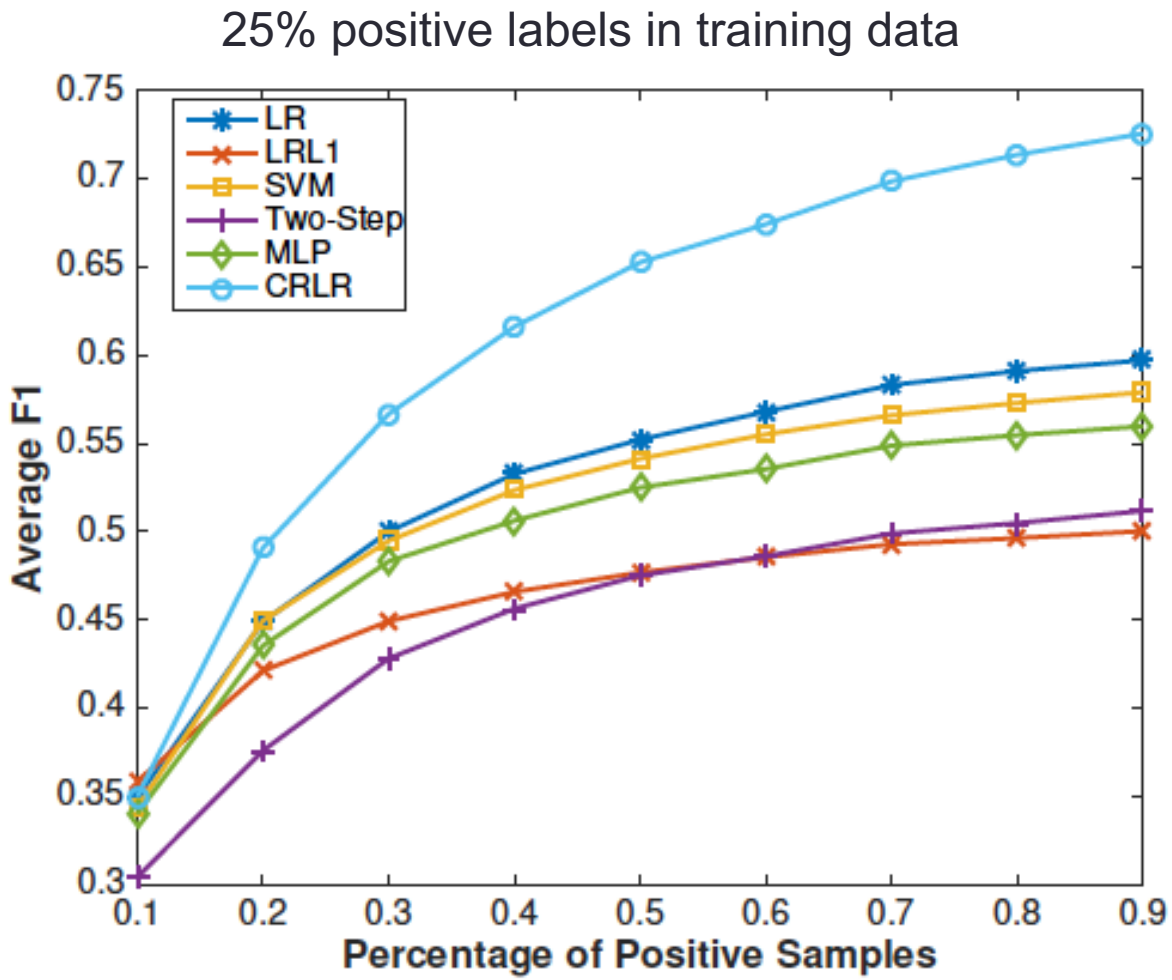
Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

Experimental Result – moderate bias



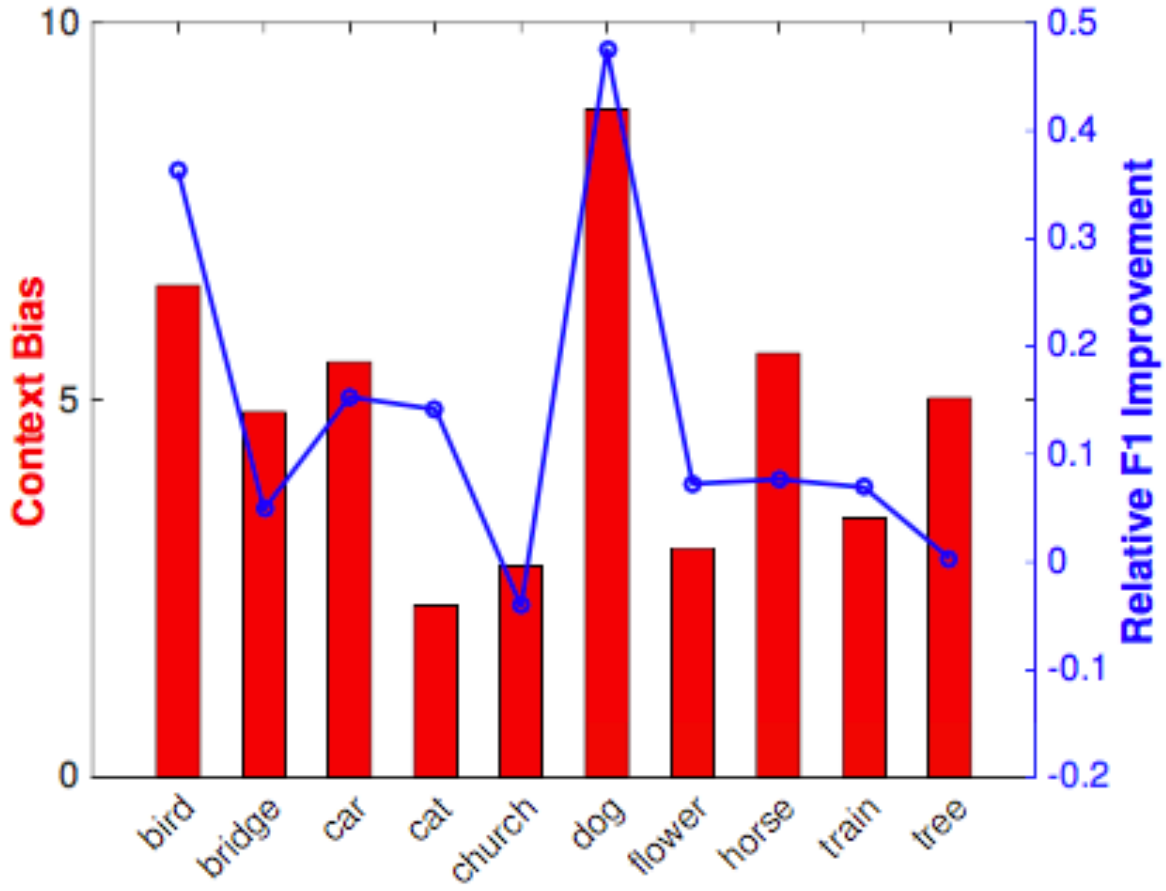
Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

Experimental Result – label bias



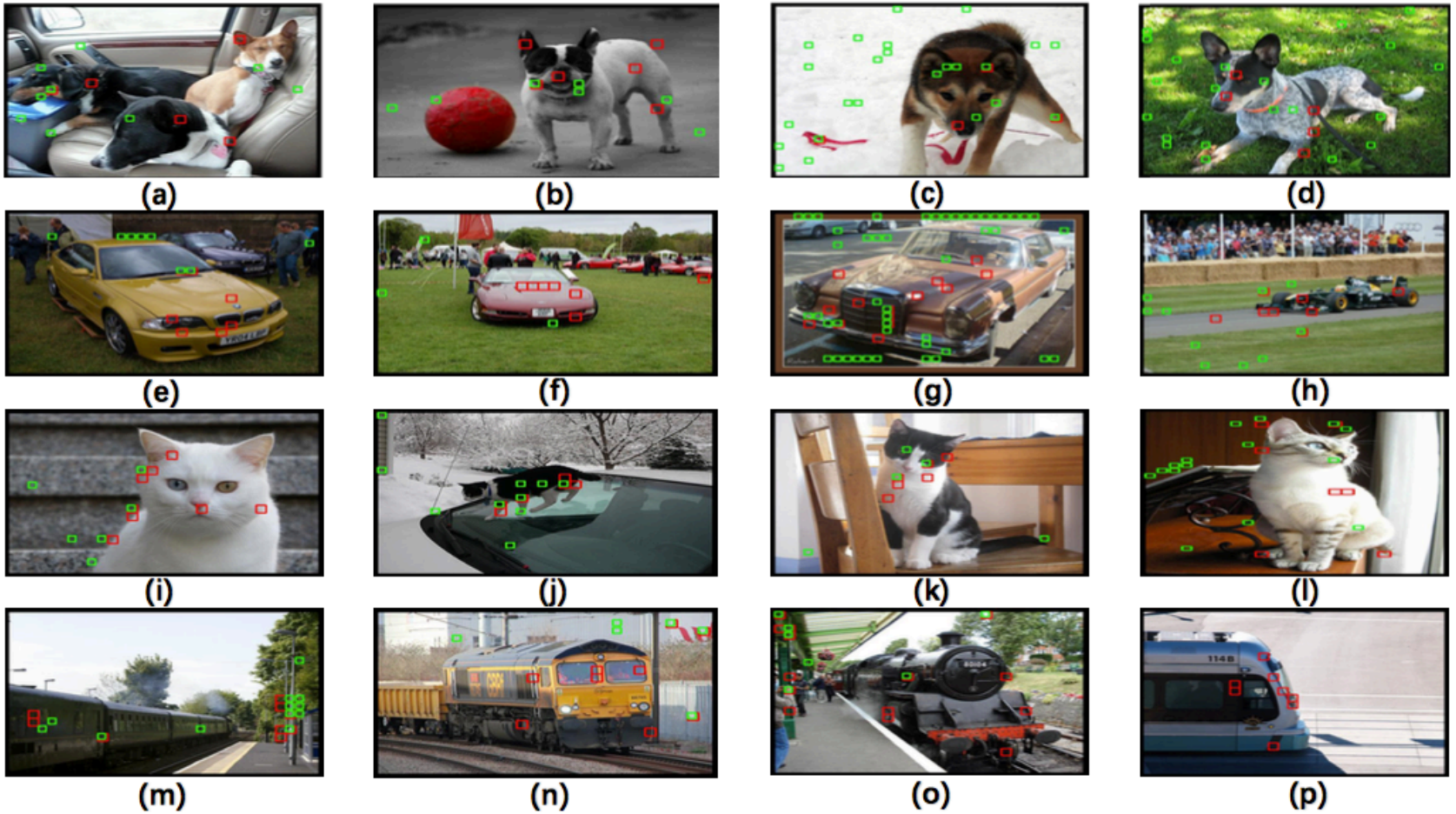
Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

Experimental Result - insights



Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

Experimental Result - insights



Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? http://arxiv.org/abs/1708.06656

Learning to Learn Image Classifiers

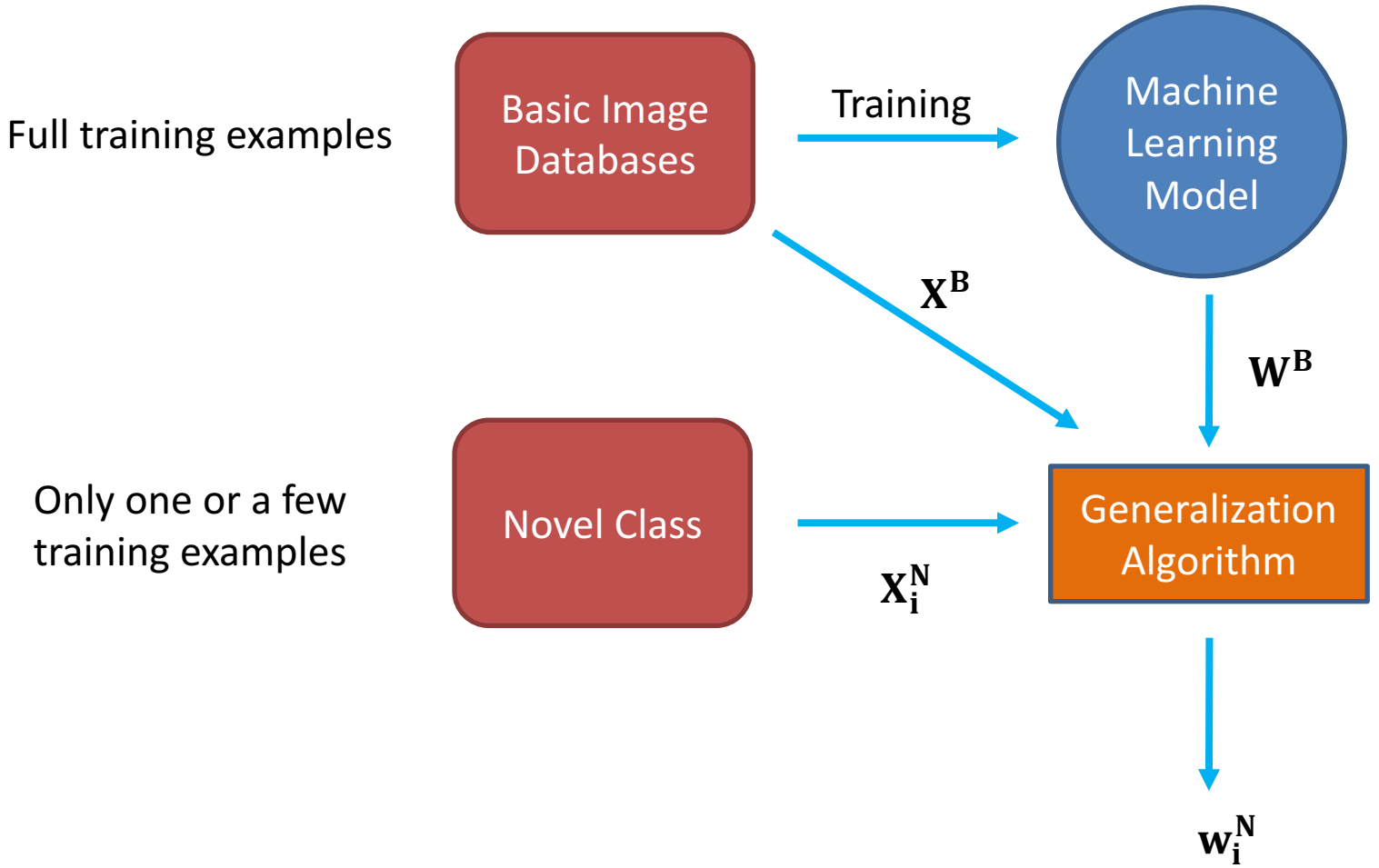
Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Problem Definition

PROBLEM 1 (LEARNING TO LEARN IMAGE CLASSIFIERS). ***Given*** *the image features of base classes* $\mathbf{X}^B$*, the well-trained base classifier parameters* $\mathbf{W}^B$*, and the image features of a novel class* i $\mathbf{X}_i^N$ *with only a few positive samples,* ***learn*** *the classification parameters* $\mathbf{w}_i^N$ *for the novel class, so that the learned classifier* $f(\cdot; \mathbf{w}_i^N | \mathbf{X}^B, \mathbf{W}^B, \mathbf{X}_i^N)$ *can precisely predict labels for the* i^{th} *novel class.*

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

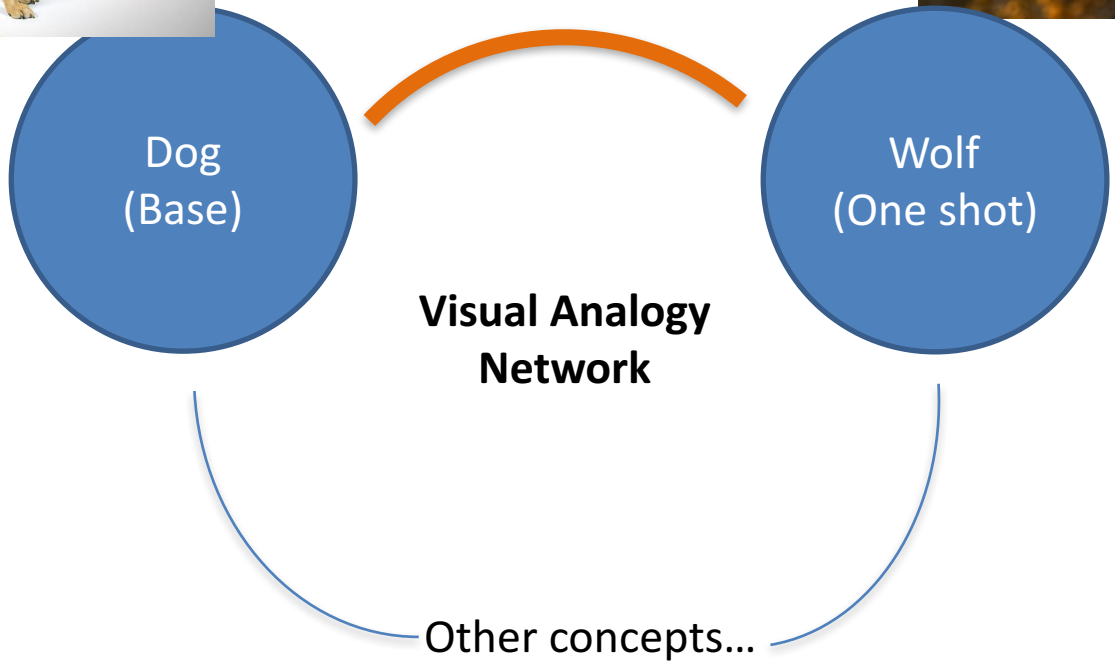
Problem Definition



Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Algorithm –VANER (Intuition)

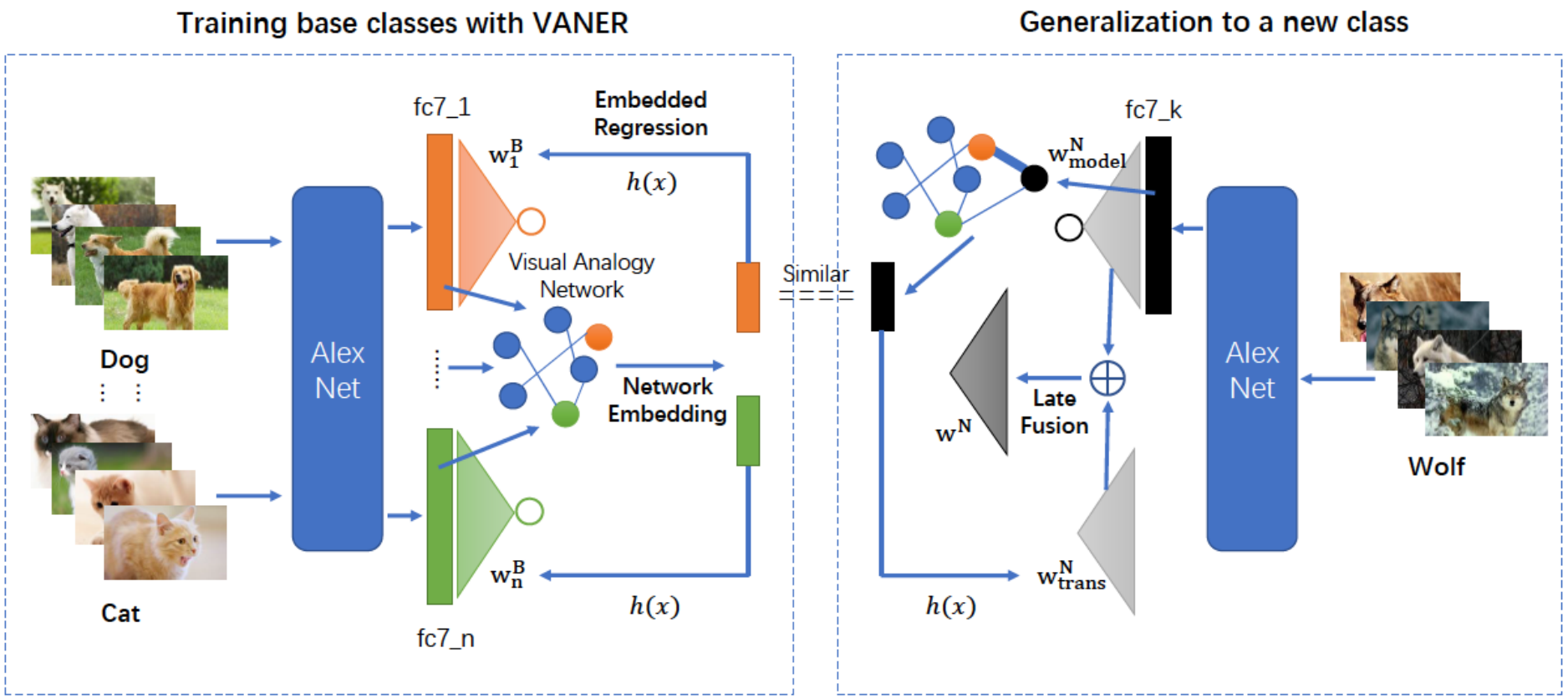
How do human learn a concept without seeing many photos?



Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

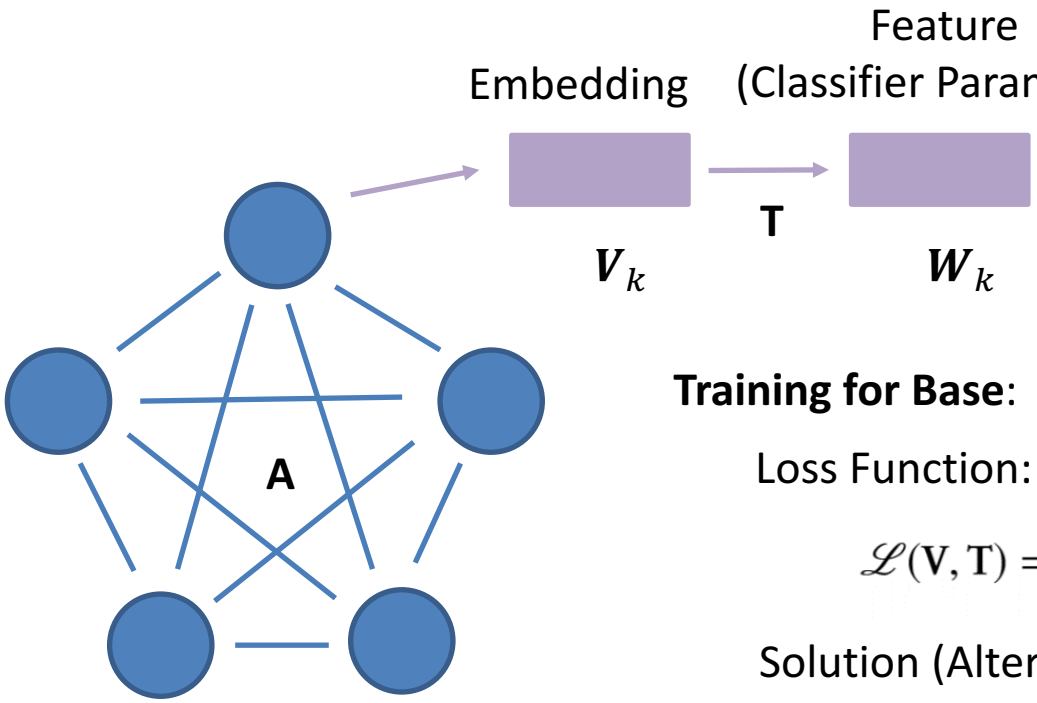
Algorithm – VANER

VANER: Visual Analogy Network Embedded Regression



Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Algorithm – VANER (Details)



A: Similarity Matrix

Training for Base:

Loss Function:

$$\mathscr{L}(\mathbf{V},\mathbf{T}) = \|\mathbf{VT}-\mathbf{W}\|_F^2 + \lambda\|\mathbf{A}-\mathbf{VV}^\top\|_F^2.$$

Solution (Alternative Coordinate Descent):

$$\begin{cases} \dfrac{\partial \mathscr{L}(\mathbf{V},\mathbf{T})}{\partial \mathbf{V}} = 2(\mathbf{VT}-\mathbf{W})\mathbf{T}^\top + \lambda(-4\mathbf{AV}+4\mathbf{VV}^\top\mathbf{V}) \\ \dfrac{\partial \mathscr{L}(\mathbf{V},\mathbf{T})}{\partial \mathbf{T}} = 2\mathbf{V}^\top(\mathbf{VT}-\mathbf{W}). \end{cases}$$

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Algorithm – VANER (Details)

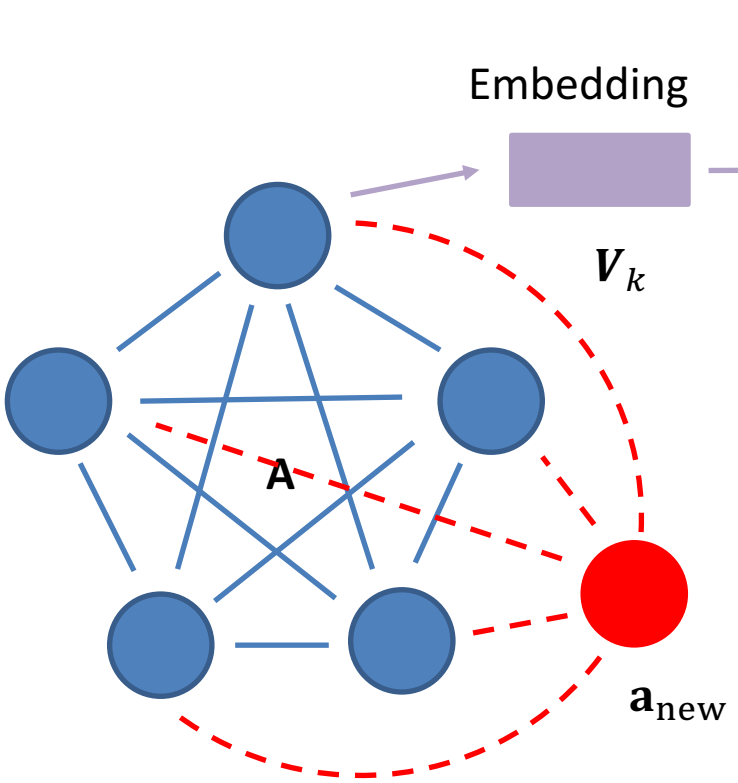
$$\mathscr{L}(\mathbf{V},\mathbf{T}) = \|\mathbf{VT} - \mathbf{W}\|_F^2 + \lambda\|\mathbf{A} - \mathbf{VV}^\top\|_F^2.$$

Keeping the precision of the predicted parameter

Keeping the structure of the visual analogy network

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Algorithm – VANER (Details)



A: Similarity Matrix

Generalizing for Oneshot:

Loss Function:

$$\mathscr{L}(\mathbf{v}_{new}) = \left\| \begin{bmatrix} \mathbf{A} & \mathbf{a}_{new}^{\top} \\ \mathbf{a}_{new} & 1 \end{bmatrix} - \begin{bmatrix} \mathbf{V} \\ \mathbf{v}_{new} \end{bmatrix} \begin{bmatrix} \mathbf{V}^{\top} & \mathbf{v}_{new}^{\top} \end{bmatrix} \right\|_F^2 .$$

Predicting parameters:

$$\mathbf{w}_{new}^{N} = \mathbf{v}_{new}\mathbf{T}.$$

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Algorithm – VANER (Details)

Decreasing the Complexity:

$$\mathscr{L}(\mathbf{v}_{new}) = \left\| \begin{bmatrix} \mathbf{A} & \mathbf{a}_{new}^{\top} \\ \mathbf{a}_{new} & 1 \end{bmatrix} - \begin{bmatrix} \mathbf{V} \\ \mathbf{v}_{new} \end{bmatrix} \begin{bmatrix} \mathbf{V}^{\top} & \mathbf{v}_{new}^{\top} \end{bmatrix} \right\|_F^2 .$$

$$\Leftrightarrow \quad \mathscr{L}(\mathbf{v}_{new}) = 2 \left\| \mathbf{a}_{new} - \mathbf{v}_{new}\mathbf{V}^{\top} \right\|_2^2 + (\mathbf{v}_{new}\mathbf{v}_{new}^{\top} - 1).$$

To speed up, eliminate the second term of the loss function:

$$\Leftrightarrow \quad \mathbf{v}_{new} = \mathbf{a}_{new}(\mathbf{V}^{\top})^{+},$$

So, we could pre-compute $(\mathbf{V}^{\mathbf{T}})^{+}$, where + represents pseudo-inverse.

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Algorithm – VANER (Late Fusion)

Initializing (w_{trans} as initialization):

$$\mathscr{L}(\mathbf{w}^N) = \left\{ \sum_{\mathbf{x} \in \mathrm{X}_T} L(f(\mathbf{x}, \mathbf{w}^N), y) \right\} + \lambda \cdot R(\mathbf{w}^N), \qquad (8)$$

Tuning:

$$\mathscr{L}(\mathbf{w}^N) = \left\{ \sum_{\mathbf{x} \in \mathrm{X}_T} L(f(\mathbf{x}, \mathbf{w}^N), y) \right\} + \lambda \cdot \left\| \mathbf{w}^N - \mathbf{w}^N_{trans} \right\|_F^2 . \qquad (9)$$

Voting (Best):

$$\mathbf{w}^N = \mathbf{w}^N_{trans} + \lambda \cdot \mathbf{w}^N_{model} . \qquad (10)$$

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Experiment Settings

- Dataset: ILSVRC 2015
- 800 Base Classes in ImageNet for training VANER, the base deep network we use is AlexNet
- 200 Novel Classes, each used for binary classification with whole base classes
- For each k-shot problem, we do 10 repeated tests with randomly split in novel class and take the average result.
- Evaluation Metric: AUC / F1 score

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Experiment Baseline

- Logistic Regression (LR)
- Weighted Logistic Regression (Weighted-LR)
- Model Regression Network (MRN)
- VANER
- VANER (-Mapping)
- VANER (-Embedding)

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Experimental Results (1) – Late Fusion

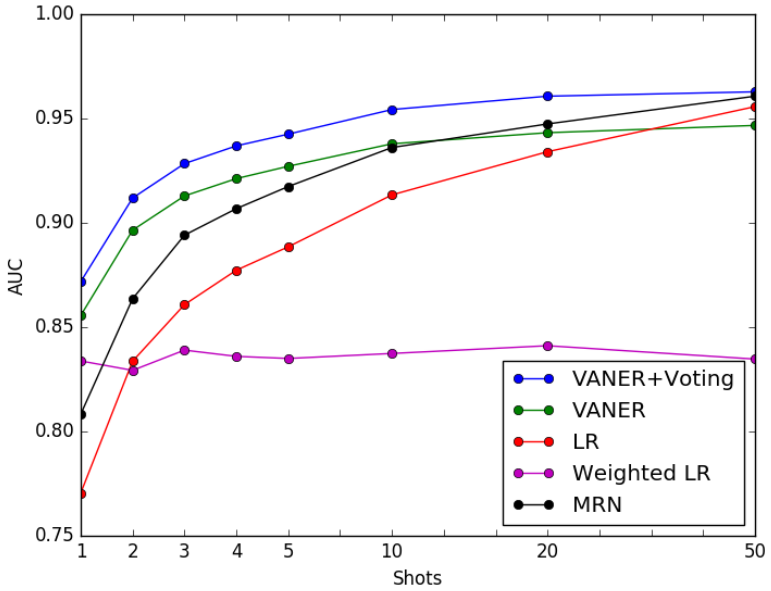
Table 2: Performance of different late fusion mechanism for k-shot problem

Algorithm	1-shot		5-shot		10-shot		20-shot	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
VANER	0.8556	0.5292	0.9271	0.6491	0.9379	0.6721	0.9432	0.6850
VANER + Initializing	0.7662	0.3941	0.9030	0.6185	0.9338	0.6887	0.9461	0.7237
VANER + Tuning	0.7923	0.4244	0.9098	0.6307	0.9365	0.7012	0.9466	0.7268
VANER + Voting	**0.8718**	**0.5671**	**0.9425**	**0.7039**	**0.9543**	**0.7343**	**0.9607**	**0.7510**

The Voting method is proved to be a better method!

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Experimental Results – Algorithm Performance



Compared with logistic regression, we can save 4/5 samples to get similar performance.

Table 1: Performance of different algorithms for k-shot problem

Algorithm	Model Transfer	1-shot		5-shot		10-shot		20-shot	
		AUC	F1	AUC	F1	AUC	F1	AUC	F1
*VANER + Voting**	Y	**0.8718**	**0.5671**	**0.9425**	**0.7039**	**0.9543**	**0.7343**	**0.9607**	**0.7510**
*VANER**	Y	0.8556	0.5292	0.9271	0.6491	0.9379	0.6721	0.9432	0.6850
VANER(−Mapping)	Y	0.8261	0.4551	0.8526	0.4807	0.8726	0.5179	0.8897	0.5394
VANER(−Embedding)	Y	0.7922	0.4335	0.9032	0.6015	0.9183	0.6347	0.9393	0.6788
LR	N	0.7705	0.3994	0.8885	0.5882	0.9134	0.6421	0.9341	0.6877
Weighted − LR	Y	0.8338	0.4680	0.8350	0.4691	0.8374	0.4711	0.8411	0.4726
MRN	Y	0.8083	0.4511	0.9175	0.6653	0.9361	0.7133	0.9474	0.7388

Experimental Results – Insightful Analysis

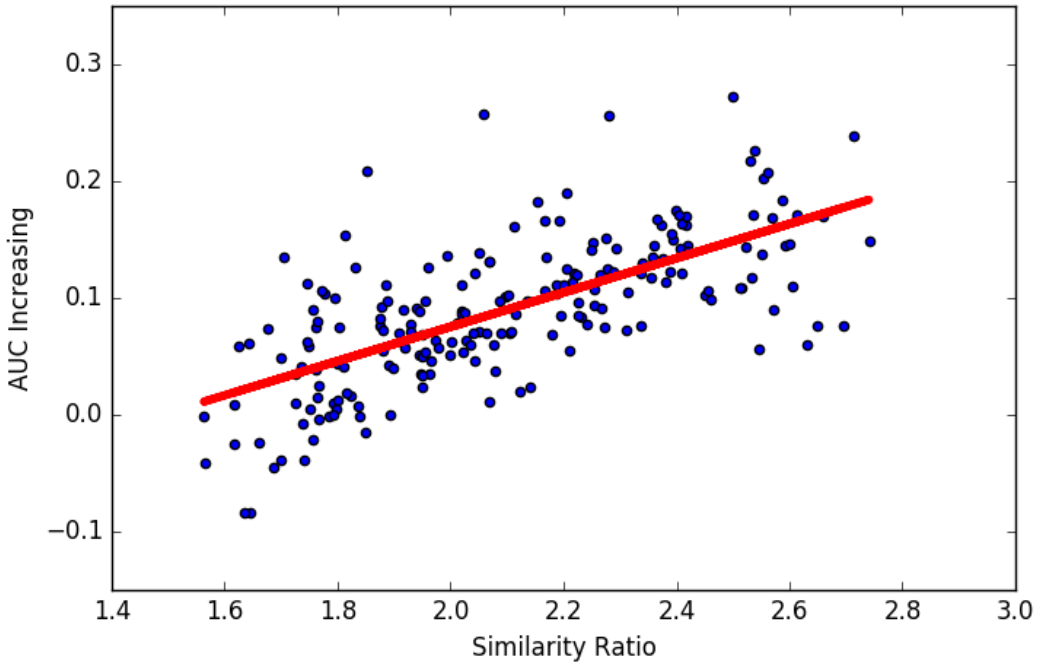
Category	LR (No Transfer)	VANER (Transfer)
Jeep	0.8034	**0.9469**
Zebra	0.8472	**0.9393**
Hen	0.7763	**0.8398**
Lemon	0.6854	**0.9583**
Bubble	**0.7455**	0.7041
Pineapple	0.7364	**0.8623**
Lion	0.8305	**0.9372**
Screen	0.7801	**0.9056**
Drum	0.6510	**0.6995**
Restaurant	0.7806	**0.8787**

Compared with no-transfer algorithm, our VANER is obviously better. However, there are some failure cases like Bubble.

What is the driving factor that controls the success of generalization?

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Experimental Results – Insightful Analysis



$$Def: \quad Similarity\ Ratio = \frac{Average\ Top-k\ Base\ similarity}{Average\ Total\ Base\ similarity}$$

$$AUC\ Increasing = AUC\ for\ VANER - AUC\ for\ LR$$

Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

Experimental Results – Embedding Similarity

The embedding layer is explainable:

Novel Class	 Jeep	 Zebra	 Lemon	 Lion	 Screen	 Restaurant
Top-3 Similar Base Classes	 Pickup  Beach_wagon  Tow_truck	 Echidna  Leopard  Cheetah	 Orange  Acorn  Granny_Smith	 Cougar  Dingo  Lynx	 Monitor  Laptop  Television	 Shoe_shop  Marimba  Bakery

Figure 3: Top-3 most similar base classes to novel class on embedding layer in 5-shot setting.

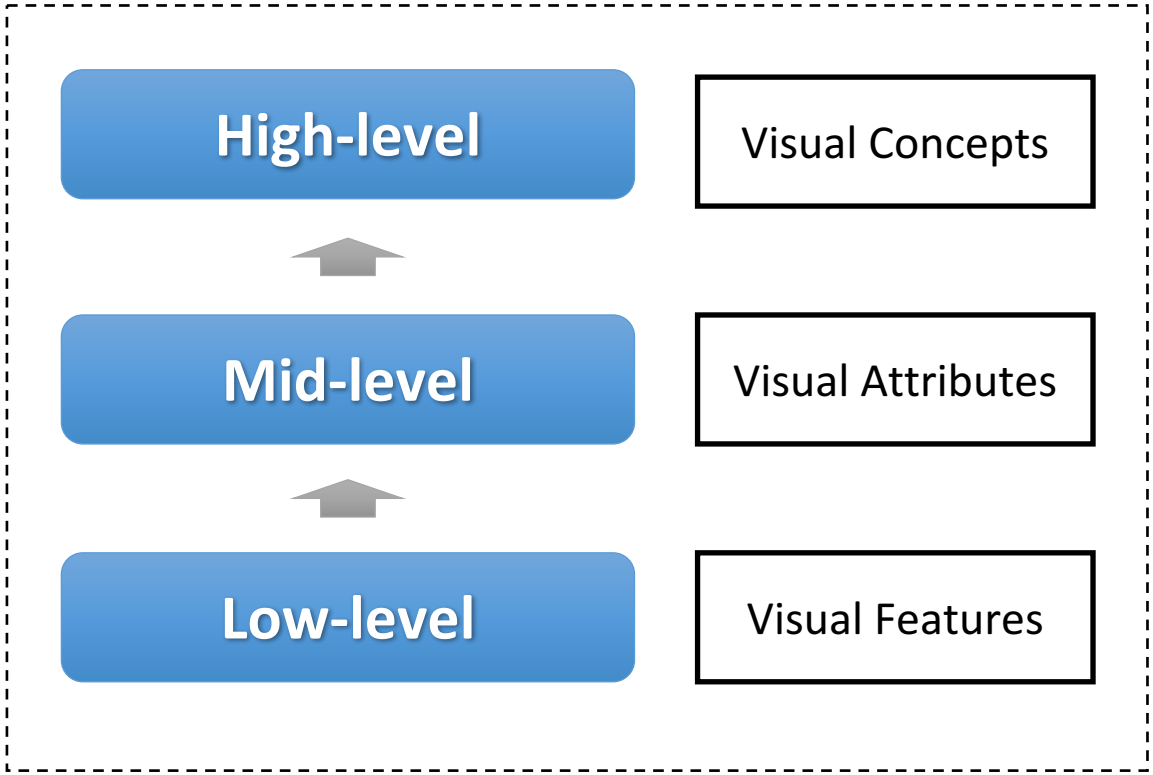
Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.0617

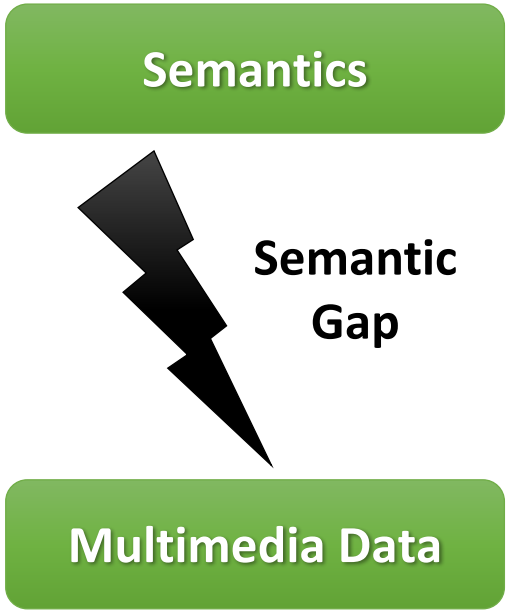
Human-like Visual Reasoning and Learning

For human:

Can we infer how people behave with visual objects?

Representation is a fundamental problem.



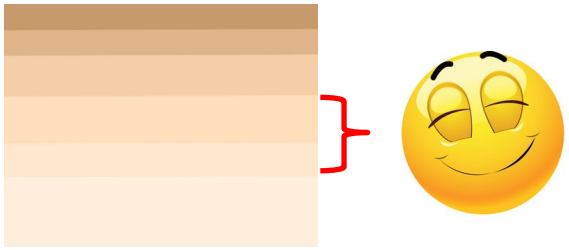


Problems of semantic-oriented representations

How much content can be described by textual semantics?

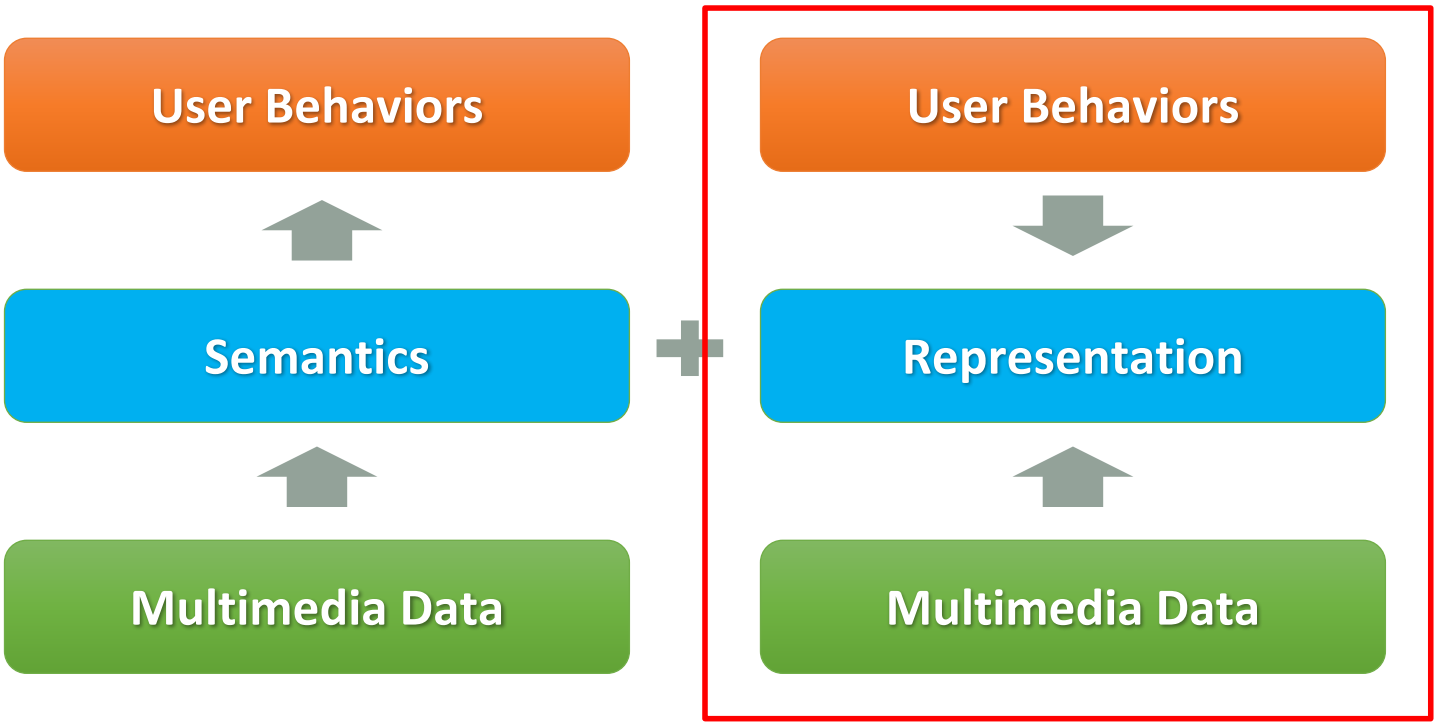
Problems of semantic-oriented representations

Are human intentions purely determined by semantics?



Revisit the Representation Learning for Multimedia

The transformation from multimedia data to semantics is lossy, and the lost information is non-trivial for inferring user behaviors.



Distance Metric: Behavioral Similarity

Behavioral Similarity

If image A and B are similar, then users should have similar behaviors on them.

Image Distance

Content Similarity

Semantic Similarity

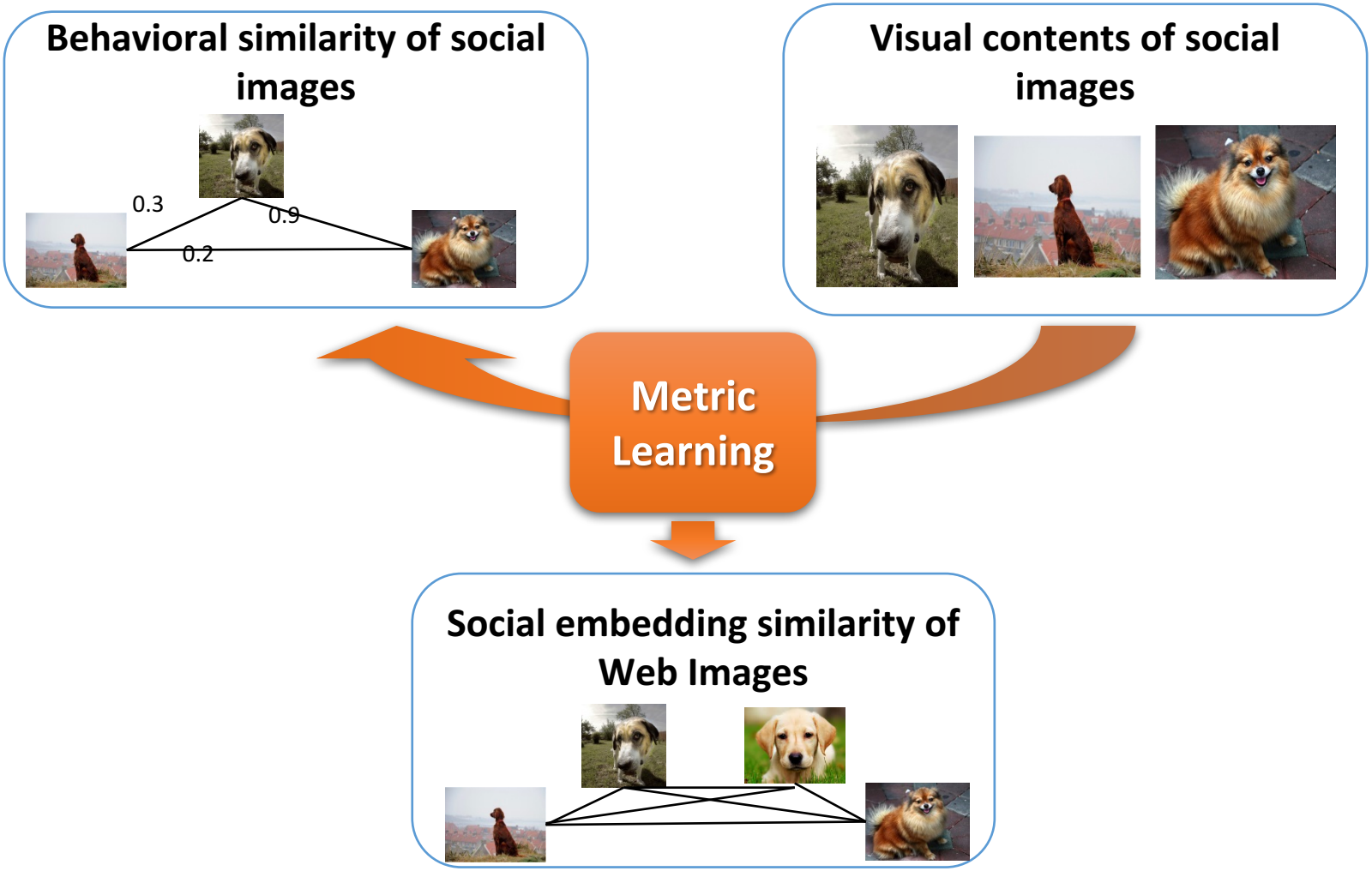
23 common users

√

0 common users

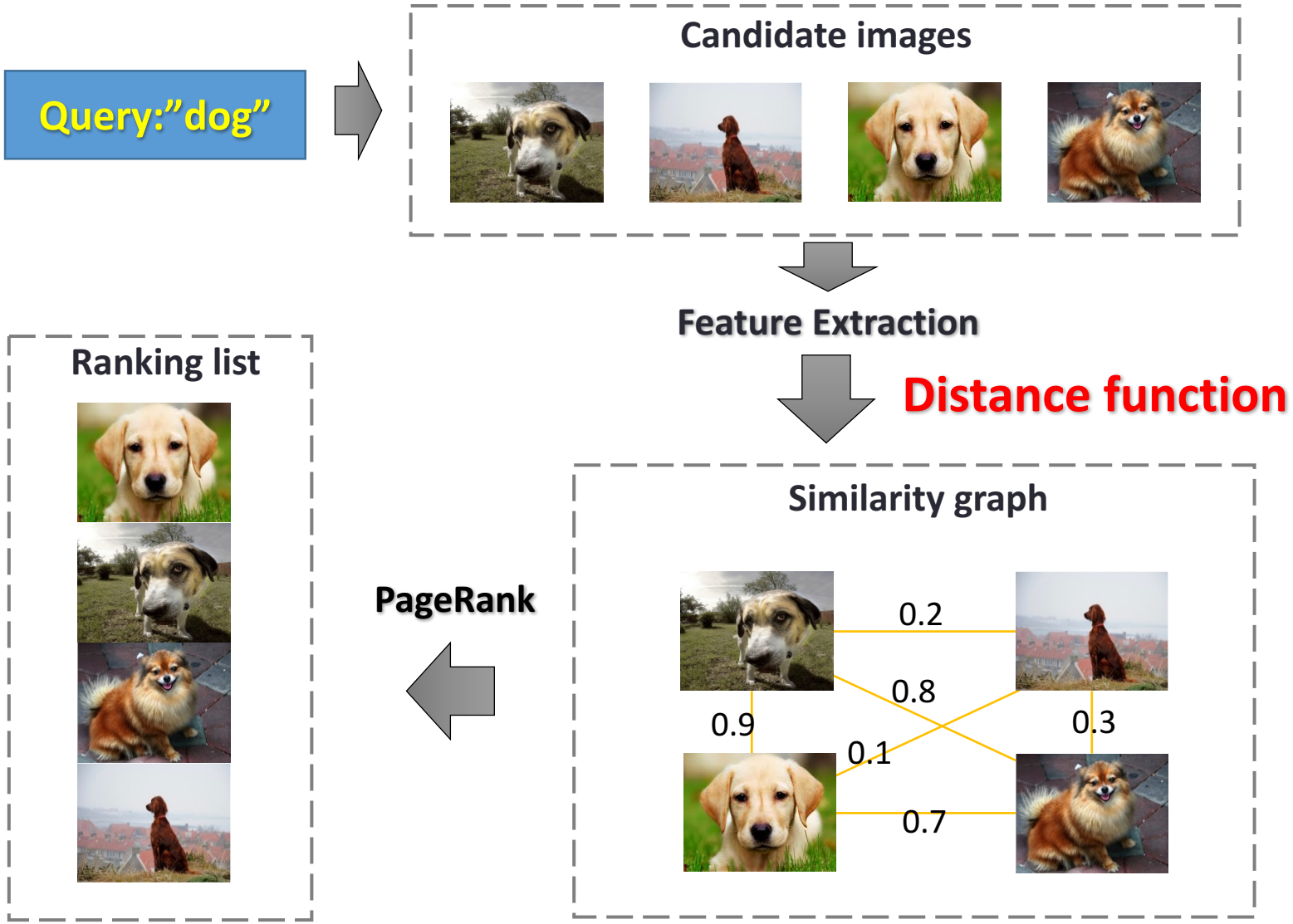
x

Old feature space, New distance metric

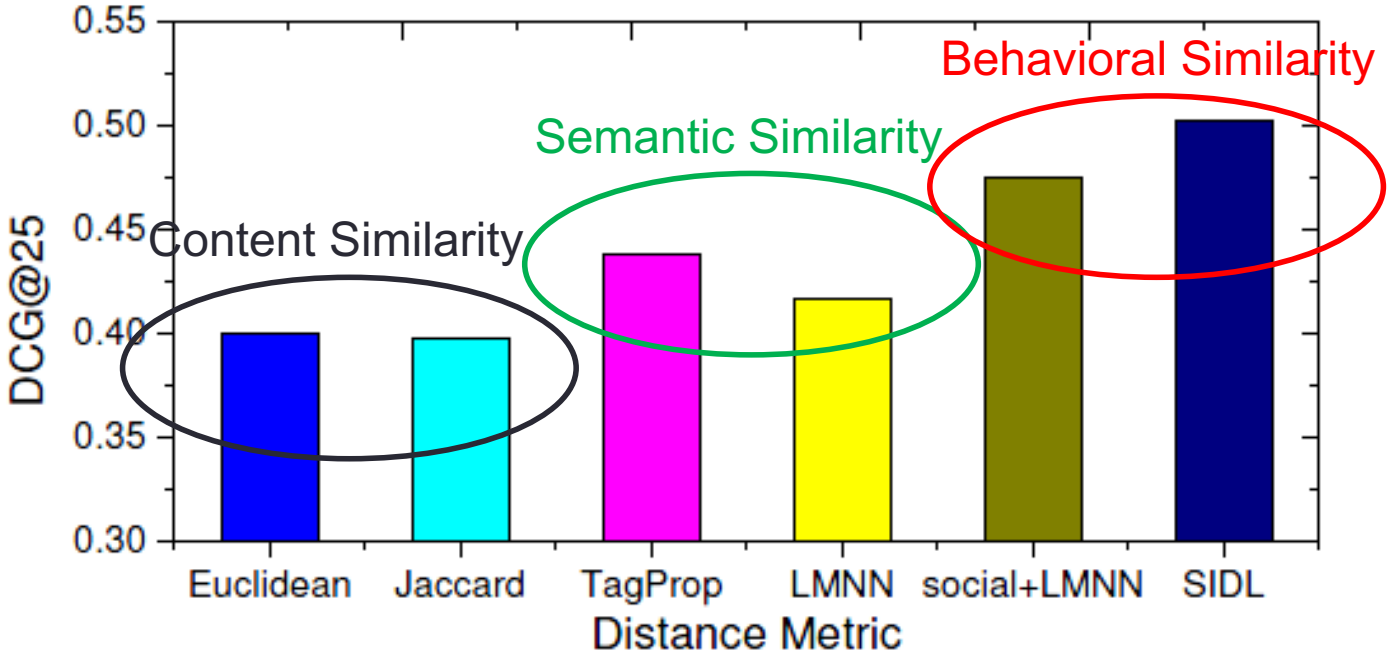


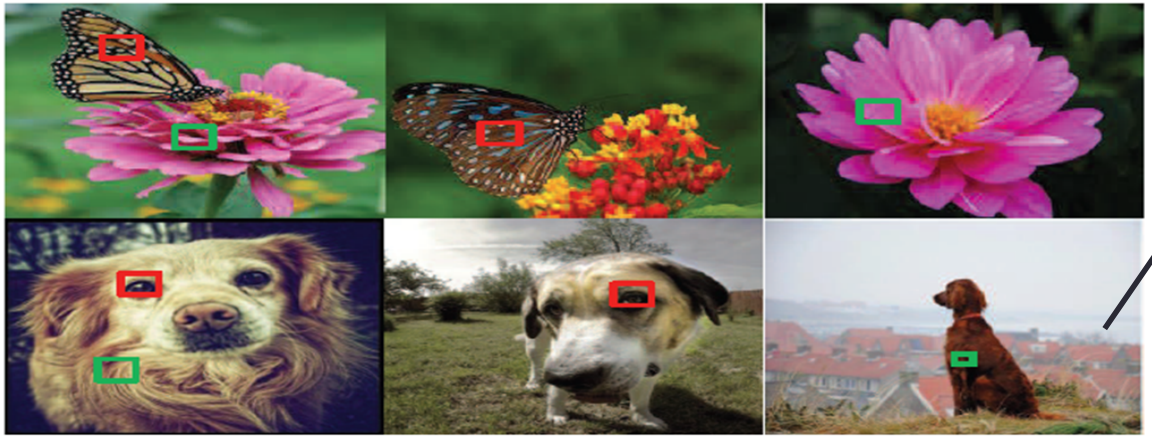
Shaowei Liu, Peng Cui, Wenwu Zhu, Shiqiang Yang, Qi Tian. Social Embedding Image Distance Learning. *ACM Multimedia*, 2014.

Learned from Flickr, applied into Bing image



Results and Insights



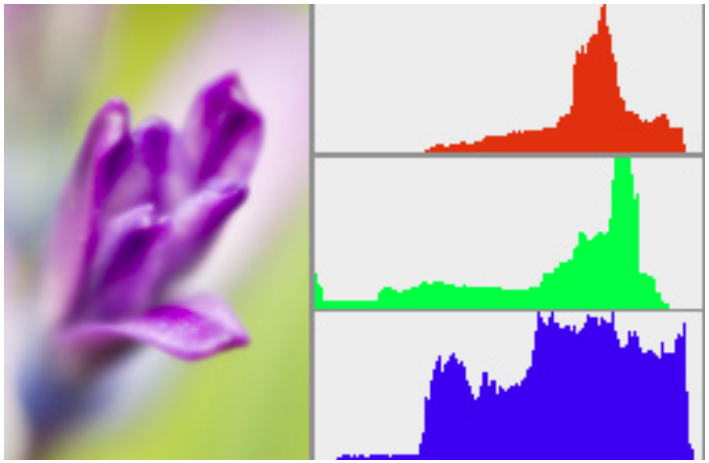


□ : high weights □ : low weights

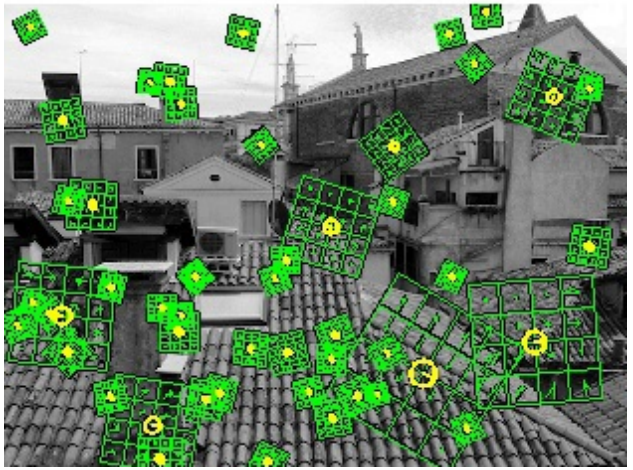
Consistent with psychovisual findings

ACM MM'14

One step further:
Can hand-crafted features well capture user intentions?



color histogram



SIFT descriptor

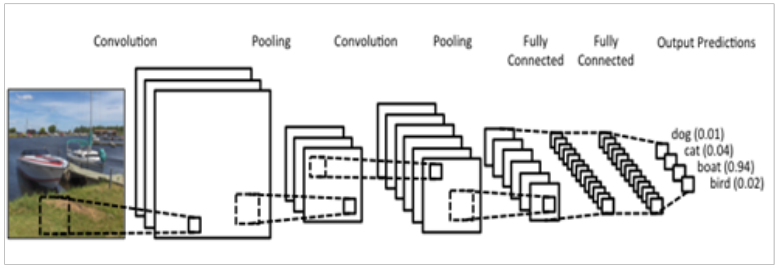
They are designed for semantics, rather than intentions.

Shaowei Liu, Peng Cui, Wenwu Zhu, Shiqiang Yang. Learning Socially Embedded Visual Representation from Scratch. *ACM Multimedia*, 2015.

Learning socially embedded image representations from Scratch

Image pixels as input

image content



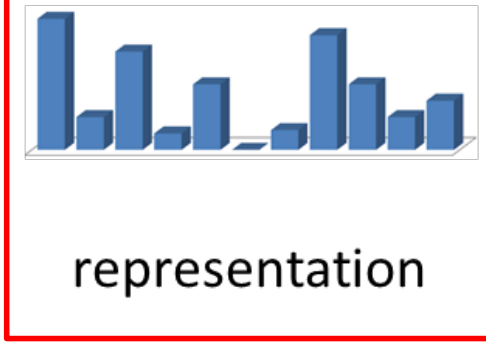
Convolutional Neural Network

Behaviors as supervision

user behavior

like

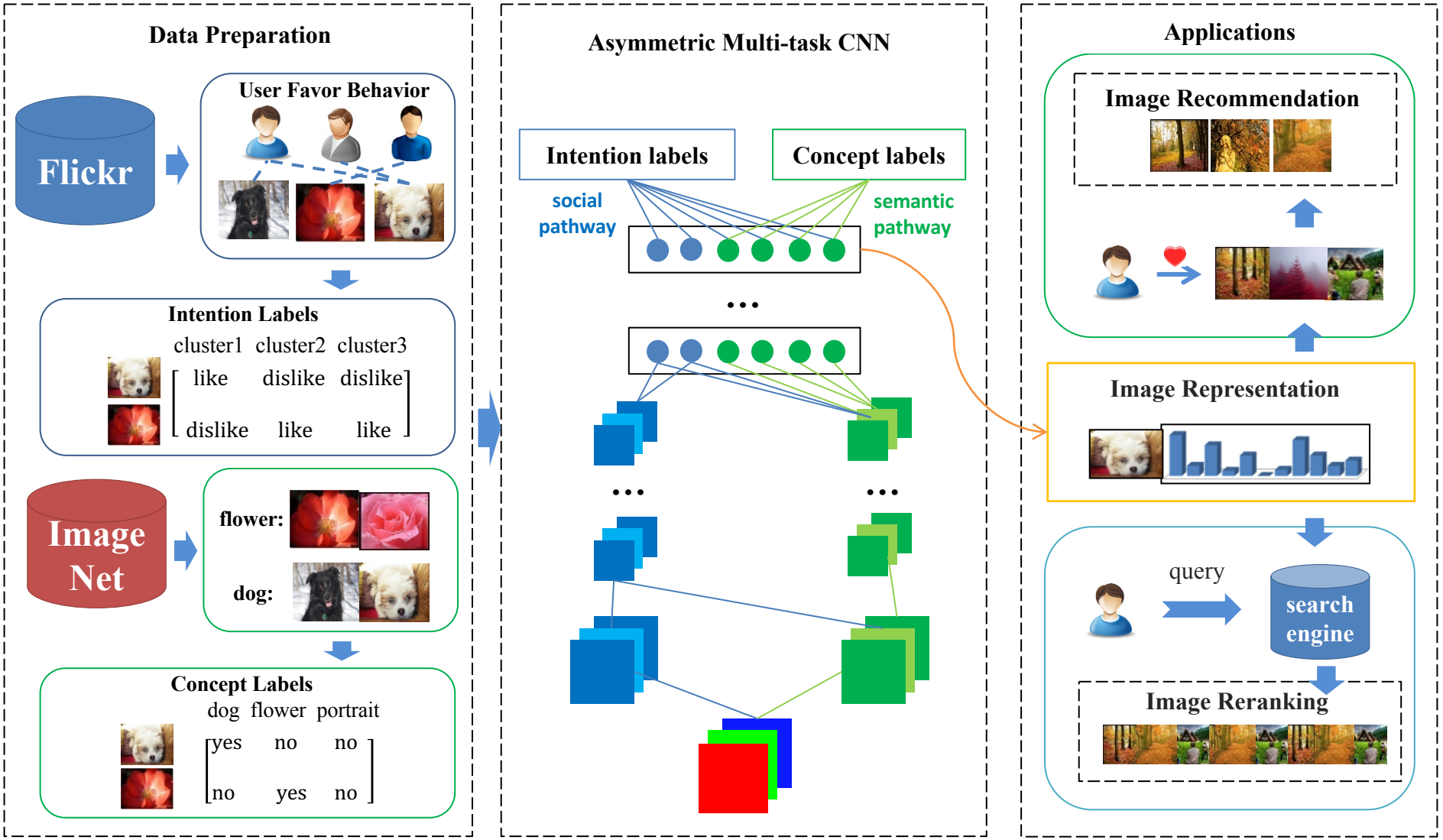
dislike



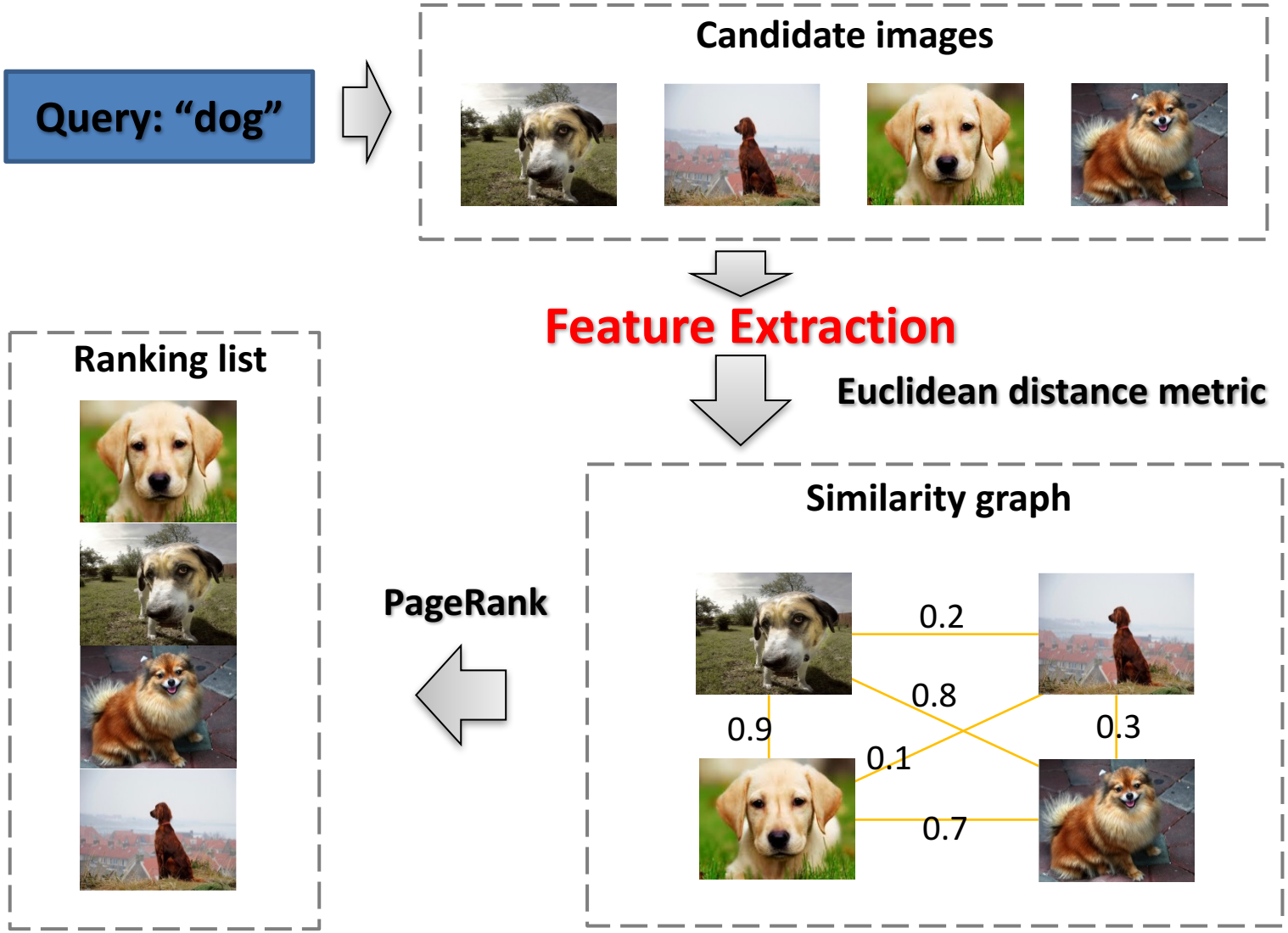
representation

Shaowei Liu, Peng Cui, Wenwu Zhu, Shiqiang Yang. Learning Socially Embedded Visual Representation from Scratch. *ACM Multimedia*, 2015.

The representation learning framework

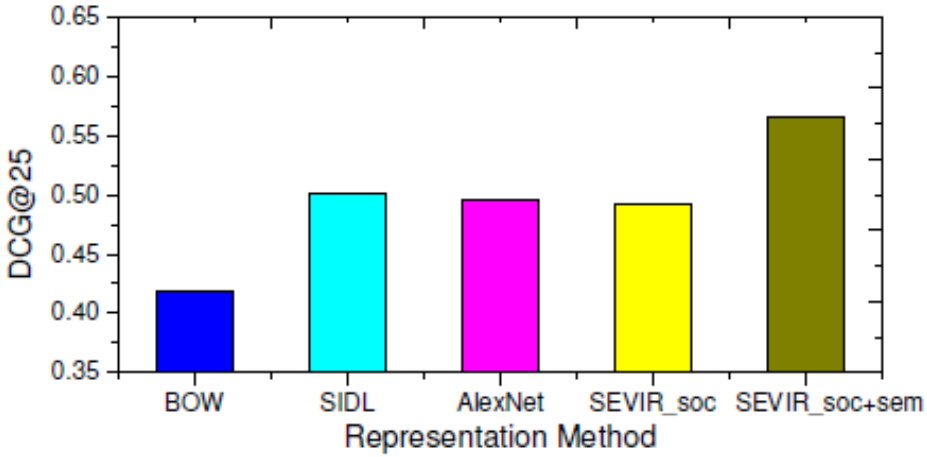


Learned from Flickr, applied into Bing image

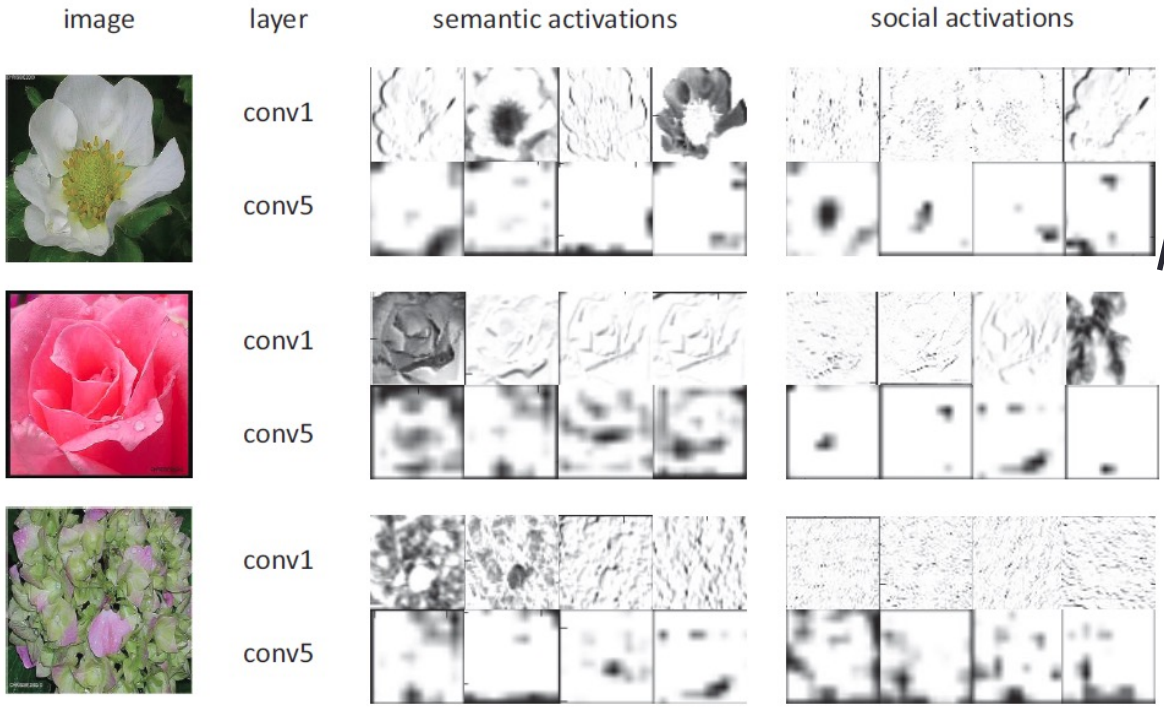


Results and Insights

ACM MM'15



- ***BOW***: handcrafted features
- ***AlexNet***: semantic oriented feature learning
- ***SEVIR_soc***: behavior oriented feature learning
- ***SEVIR_soc+sem***: behavior and semantic oriented feature learning



The social path capture different aspects from semantic path. How to interpret?

Summary and Messages

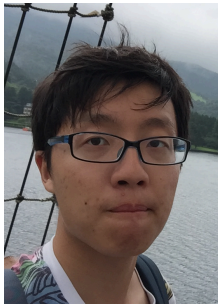
- Beyond parameter tuning, it is more important to think about the learning mechanism.
- Human-like learning and reasoning is the valuable source to get inspirations.
- From black-box prediction models to explainable learning and reasoning processes is more meaningful.
- Learning comprehensive and interpretable representations for multimedia to reflect user behaviors.

References

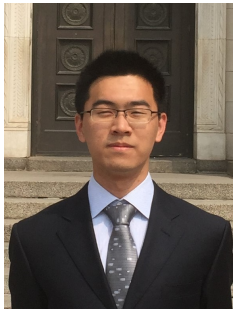
- Peng Cui, Wenwu Zhu, Tat-Seng Chua, Ramesh Jain. Social-Sensed Multimedia Computing. *IEEE Multimedia*, 2016.
- Zheyan Shen, Peng Cui, Kun Kuang, Bo Li. On Image Classification: Correlation V.S. Causality? *http://arxiv.org/abs/1708.06656*
- Linjun Zhou, Peng Cui, Shiqiang Yang, Wenwu Zhu, Qi Tian. Learning to Learn Image Classifiers with Informative Visual Analogy, https://arxiv.org/abs/1710.06177
- Shaowei Liu, Peng Cui, Wenwu Zhu, Shiqiang Yang. Learning Socially Embedded Visual Representation from Scratch. *ACM Multimedia*, 2015.
- Shaowei Liu, Peng Cui, Wenwu Zhu, Shiqiang Yang, Qi Tian. Social Embedding Image Distance Learning. *ACM Multimedia*, 2014.
- Kun Kuang, Peng Cui, Bo Li, Meng Jiang, Shiqiang Yang. Estimating Treatment Effect in the Wild via Differentiated Confounder Balancing. *KDD*, 2017.
- Kun Kuang, Peng Cui, Bo Li, Meng Jiang, Fei Wang, Shiqiang Yang. Treatment Effect Estimation with Data-Driven Variable Decomposition. *AAAI*, 2017.

Acknowledgement

Students



Zheyan Shen
Tsinghua U



Linjun Zhou
Tsinghua U

Kun Kuang
Tsinghua U

Shaowei Liu
Tsinghua U

Collaborators

Ramesh Jain
UCI

Tat-Seng Chua
NUS

Qi Tian
UTSA

Thanks!

Peng Cui
cuip@tsinghua.edu.cn
http://media.cs.tsinghua.edu.cn/~multimedia/cuipeng/