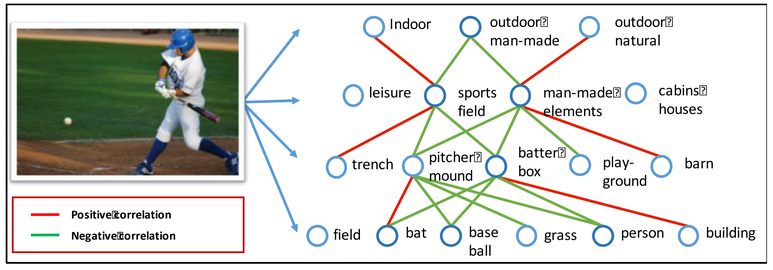
Learning Structured Inference Neural Networks with Label Relations

Hexiang Hu[†‡] Guang-Tong Zhou[†] Zhiwei Deng[†] Zicheng Liao[‡] Greg Mori[†]

[†]Simon Fraser University [‡]Zhejiang University

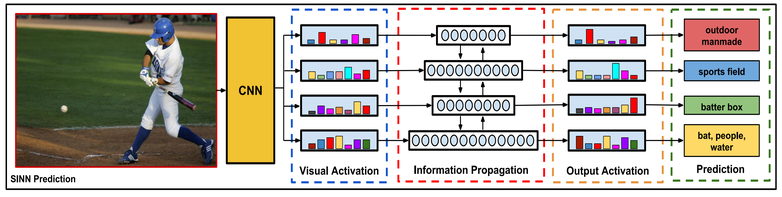
IEEE Conference on Computer Vision and Pattern Recognition

Overview



Motivation

- **Fact**: Image labels are from various *concept layers*.
  - **Coarse-grained labels** depicts high level abstraction
  - **Fine-grained labels** describes of major components
  - **Attribute labels** reveals properties
- **Idea**: Explore *label relations* among concept layers.
  - **Joint inference**: better label prediction by leveraging structured label relations
  - **Partial inference**: better prediction of fine-grained labels if given coarse labels



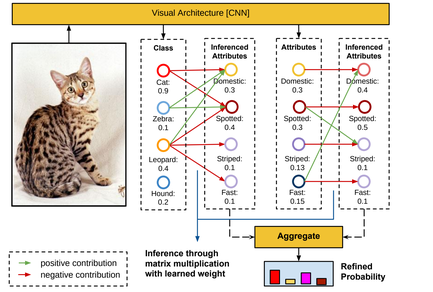
Contributions

- **An inference model**: to leverage label relations among concept layers.
  - **Joint & partial inference**: designed to predict labels at any concept layer.
  - **Label propagation**: pass context within & between concept layers.
- **An end-to-end learning framework**: CNN for images + RNN for labels.
- **A reverse sigmoid module:** to smooth flow among observed labels and activations.

Model Details

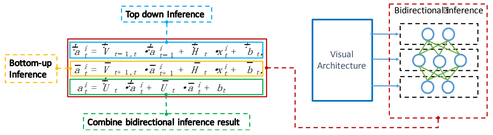
Label Propagation in Neural Networks

- **Concept layer** ➔ recurrent unit in RNN
  - Capture label relations among concept layers
- **Untying recurrent weights**
  - Different concept layers have different labels



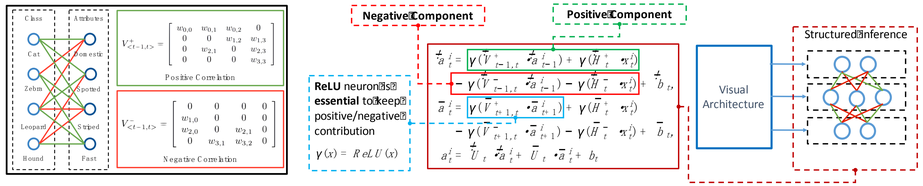
Bidirectional Inference Neural Network

- Propagate information **bi-directionally**
- **Top-down**: from coarse to fine-grained labels
- **Bottom-up**: from fine-grained to coarse labels



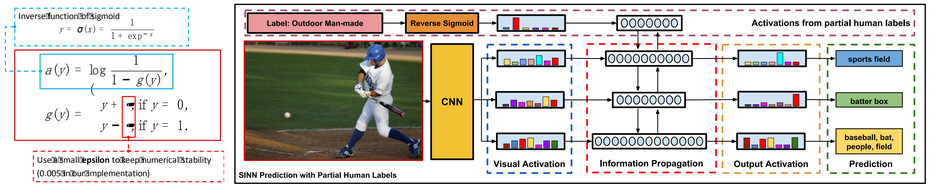
Structured Inference Neural Network

- Pre-defined **label relation matrix**: to constraint information propagation paths
- **ReLU module**: to explicitly enforce positive and negative label relations



Inference with Partially Observed Labels

- How to feed observed labels into network? **Reverse Sigmoid Module**
- Enable smooth tensor flow between observed labels and RNN activations.



Experiments

Datasets

Animal with Attributes (AwA) [Lampert et al. 2009]

Label Organization

28 taxonomy terms;
50 animal classes;
85 attributes;

Ex1: predict entire label set

- Taxonomy terms are constructed from WordNet as [J. sung et al. 2012]
- Knowledge graph constructed by combining class-attributes graph with flat taxonomy graph

NUS-WIDE [Chua et al. 2009]

Label Organization

698 image groups;
81 concepts;
1k tags;

Ex2: predict 81 concepts with observing tags/image groups

- Knowledge graph produced by WordNet using semantic similarity
- 698 image groups are constructed from image meta data

SUN 397 [Xiao et al. 2012]

Label Organization

3 coarse scenes;
16 general scenes;
397 fine-grained scene categories;

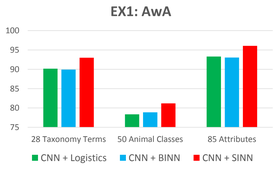
Ex 1: predict entire label set

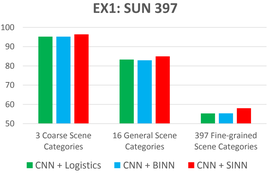
Ex 2: predict 397 fine-grained scene categories with observing 3 coarse scene category

- Knowledge graph provided by dataset

Experiment 1: Inference on All Concept Layers

- Predict labels on all concept layers. Measured by mean Average Precision (**mAP**).



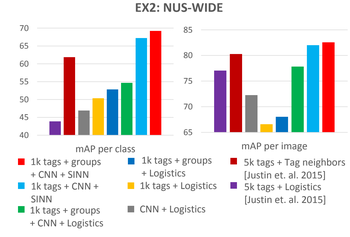


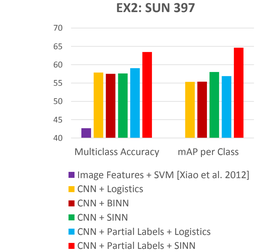
CNN + Logistic: campus
Observed Label: outdoor/man-made
Our Predictions: abbey
Ground-Truth: abbey

Experiment 2: Inference with Partially Observed Labels

Predict labels of some concept layers given labels on the other layers

- **NUS-WIDE**: Predict **81 concepts** with observation on **1k tags**
- **SUN 397**: Predict **397 fine-grained scene categories** with observation on **3 coarse scene categories** (visualization ☞)





CNN + Logistic: building facade
Observed Label: outdoor/man-made
Our Predictions: library/outdoor
Ground-Truth: library/outdoor

CNN + Logistic: patio
Observed Label: outdoor/natural outdoor/man-made
Our Predictions: picnic area
Ground-Truth: picnic area

SFU Vision and Media Lab