DEVELOPING INTERFACES FOR LABELING OBJECT RELATIONSHIPS IN IMAGES USING CROWDSOURCING

Rajan Vaish, Justin Johnson, Michael Bernstein, James Davis
rvaish@cs.ucsc.edu
Introduction
Crowdsourcing

Who/what is crowd in crowdsourcing?
Crowdsourcing

- Who/what is crowd in crowdsourcing?

It’s us.
We make the crowd.
We cause the crowdsourcing to occur.
Why crowdsourcing?

- Crowdsourcing is powerful.
- Human intelligence is invaluable and unmatchable.
- To accomplish worthwhile goals with global impact.

However,
Limitations of crowdsourcing

- Current crowdsourcing methods and platforms have limitations.
- Amazon’s Mechanical Turk is the largest crowdsourcing platform, with over 500K workers, in 190 countries.
Limitations of crowdsourcing

- We fall short of the broader participation base, or expert knowledge.

Crowd not from marketplaces
Examples:

- What happens when we get crowd to mobilize?
More examples...

.. and this happens...

**MONEY FROM THE CROWD**

Crowdfunding activity in the world

- **$1.5 B** raised worldwide
- **+90%** expected growth in 2012

**Canadian Top Funding**

- **Pebble Watch**
  - Castle Story: $10M
  - Alpha Watch: $320K
  - $700K

**JOBS Act overview**

- **Timeline**
  - April 2012: Signed by Obama
  - January 2013: Goes into effect
- **Rules for Start-up**
  - Can sell up to $1M of stocks a year to unlimited investors
- **Rules for Investors**
  - <$100K: max $2000/year
  - >$100K: max 10% of income

**Crowdfunding Models**

- **Donation or Reward**
- **Lending**
- **Equity**

**NEWGROUNDS**

Tasks via SMS for phone users in developing world
Research question

Is it possible to mobilize and engage people to harness their expertise, intelligence, time and potential to accomplish worthwhile complex goals?

- **Problem Domain (what):** What problem should the crowd focus on?
- **Process (how):** How should crowd execute a process to accomplish solution to the problem?
Types of crowd: terminologies used

- MTurk Crowd – crowd found on marketplaces
- Expert Crowd – crowd/people with an area of expertise
- Semi-Expert Crowd – non expert crowd/people not found on marketplaces
Overview and the Problem Domain (What)

Bike (subject)

part of (predicate)

Wheel (object)
Current State of Computer Vision

Computer vision is not there yet, but high level scene understanding can improve the state of computer vision. How can we collect such data?
Fixing computer vision!

- **Aim:** To improve the state of computer vision research by collaborating with the Stanford Visual Genome Project (experts) - to help build the largest visual knowledge graph in history.

However, what does it mean to us? and how can crowdsourcing help in this project?
Current approach of image search

Query:

Police spraying protester
Structured Search

Query: Police spraying protester
Use Visual Genome
Use Visual Genome
Use Visual Genome

- Police
- spraying
- Protester
Visual Genome Structured Results

Police spraying protester

Can help find non-captioned images

What is a Visual Knowledge Base?

- Consists of an image with labeled objects
What is a Visual Knowledge Base?

- Consists of an image with labeled objects
What is a Visual Knowledge Base?

- Objects have properties

- Van → Metal
- Chair → Plastic
- Person

- Person
- Van
- Chair
What is a Visual Knowledge Base?

- Consists of a large number of images
What is a Visual Knowledge Base?

• Consists of a large number of images
Initial Results: Objects Example

Labeled objects on 5000 images: ~172,000 objects total
Binary Relations

Manually labeled relationships on 200 images by Mturk Crowd

Bike (subject)

part of (predicate)

Wheel (object)
Building the KB: Crowdsourcing

Divide work into small subtasks: Object labeling
Building the KB: Crowdsourcing

Divide work into small subtasks: Object labeling

Worker 1:
Person
Person
Van
Building the KB: Crowdsourcing

Divide work into small subtasks: Object labeling

Worker 2:
Chair
Van
Person
Building the KB: Crowdsourcing

Divide work into small subtasks: Object labeling

Worker 3:

Building

Person

Tree
Building the KB: Crowdsourcing

Divide work into small subtasks: Object labeling

Combined:

Building
Tree
Chair
Van
Person
Person
Person
Building the KB: Crowds + Machines

Crowdsourced Annotations

Knowledge Base

Human Verification

ML Classifiers

Predicted Annotations
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:
Objects: 40 per image
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:

Objects: 40 per image

Assume 1¢ per object

$0.40 per image
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:

Objects: 40 per image = $0.40 per image
How much would this cost on Mechanical Turk?

Want:
Objects: 40 per image = $0.40 per image
Attributes: 10 per object
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:
Objects: 40 per image = $0.40 per image
Attributes: 10 per object
  400 per image
  Assume 0.5¢ per attribute
= $2 per image
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:

Objects: 40 per image = $0.40 per image
Attributes: 10 per object = $2 per image
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:
Objects: 40 per image = $0.40 per image
Attributes: 10 per object = $2 per image
Relationships: Check each object pair
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:

Objects: 40 per image = $0.40 per image
Attributes: 10 per object = $2 per image
Relationships: Check each object pair
1600 pairs per image
= $16 per image (1¢ per object pair)
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:

Objects: 40 per image = $0.40 per image
Attributes: 10 per object = $2 per image
Relationships: Check each object pair = $16 per image
How much would this cost on Mechanical Turk?

Want:
- Objects: 40 per image = $0.40 per image
- Attributes: 10 per object = $2 per image
- Relationships: Check each object pair = $16 per image

TOTAL: $18.40 per image
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

Want:

Objects: 40 per image = $0.40 per image
Attributes: 10 per object = $2 per image
Relationships: Check each object pair = $16 per image
TOTAL: $18.40 per image
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

$18.40 per image
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

$18.40 per image
1k images = $18k
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

$18.40 per image
1k images = $18k
1M images = $18M
Building the KB: Crowdsourcing

How much would this cost on Mechanical Turk?

$18.40 per image
1k images = $18k
1M images = $18M

• Very expensive at the moment! — discourages crowdsourcing images at this price.
• Tasks worth millions of dollars is waiting to get done, therefore, potentially helping people around the world to earn money.
• Can we bring the price down and create a win-win situation?
The Plan

Given the problem domain, the plan was to explore the possibility of lowering the cost of the work done by Mturk crowd - by involving expert and semi-expert crowd.
The Project Process (How)
Goal

- To create interactive user-interfaces which can facilitate Mturk crowd contributions and bring the price down for binary labeling!

- Recruit semi-expert crowd to brainstorm out-of-the box ideas and develop interfaces better than the one developed by experts at Stanford (benchmark).
Basically,

- Trying to harness the potential of semi-expert crowd to work under the mentorship of expert crowd to create an interface which can lower the cost of work done by Mturk crowd.
Project Details

- Large number of images provided by expert crowd.
- The interface should allow crowd to tag object and label semantic relationships between object instances.
- Each relationship takes the form of a triple (subject, predicate, object).

(car, on, road) (person, next to, building) (person, reading, book) (car, crashing into, tree)
Objects (part of)
Binary Relationships

- supports
- occludes
- next to
- part of
- grouped with
- above
- inside
The process — recruitment

- Presented a talk at CMPS 119, and got nine students excited about the idea.
- Students were our expert crowd, with computer science background, but not computer vision.
- The project would eventually become their class project as well — and would follow course schedule.
- Research and grades were the motivation behind their contribution.
The process – techniques

- A group of nine students followed a unique research process - including phases like brainstorming, prototyping, development and evaluation.

- To make the process efficient, expert crowd was mentored by Stanford and UCSC graduate students.

- Crowd teams were merged and rearranged depending on timeline and resource requirements.
The process flow

- Group of nine students were divided into four teams, to encourage diversity in ideas and approach.
The process flow

- Different teams went through a parallelized development cycle.
- Every week peer review process determined the best work, and all semi-expert teams would get mentorship from experts.
The process flow

- Teams were merged to speed up the development and evaluation process — and to bring common approaches and interests together.

- A final evaluation between team AB and CD was conducted, where team AB’s interface outperformed.
The process flow

- To evaluate the success of this process, interface by the team of semi-experts (AB) was to be compared against the team of experts (Stanford).
Results – interface by Stanford

**Task**

- 1 / 5 relationship phrases
- 0 / 5 description phrases
- 1 / 15 phrases

**Completed phrases**

- dog next to dog

**Write a phrase about the image:**

- dog

[Image of two dogs, one holding a stick and both labeled as "dog"]
Results – interface by expert crowd

Phase 1: Labeling an object

Phase 2: Labeling relationship
Evaluation

- Comparative user-studies were conducted between the interfaces (nine users participated).

- Users performed labeling task faster on the interface developed by the semi-expert crowd (SE) over Stanford’s expert crowd (E).
Evaluation - preference

We used Likert scale (1-6), where 1-3 for SE, and 4-6 for E.

- Quality of instructions: 88% preferred SE.
- Usability: 63% preferred SE.
- Design or Concept: 63% preferred E.
- Features: 63% preferred E.
- Enjoyment: 63% preferred SE.
Evaluation - preference

We used two Likert scale (1-6), where each scale represented SE and E.

- Overall score: SE (35) vs. E(31).

Therefore, we learned that the interface developed by semi-experts was on par with the one developed by experts – causing the cost of work done by Mturk crowd to go down.
Thank you! Questions?

rvaish@cs.ucsc.edu

This project was in part funded by the Los Alamos National Laboratory