A Poodle or a Dog ?

Evaluating Automatic Image Annotation Using Human Descriptions at Different Levels of Granularity

Josiah Wang   Fei Yan   Ahmet Aker   Robert Gaizauskas

Department of Computer Science, The University Of Sheffield.

Centre for Vision, Speech and Signal Processing, University of Surrey

Work presented as part of the Visual Sense (ViSen) project, funded by the chist-era D2K programme

Main Idea

Label this picture with an object name

Main Idea

Label this picture with an object name

Dog Poodle Mammal

Animal Pet Canine

Main Idea

System 1

1. Dog
2. Sheep
3. Sandal
4. Chihuahua
5. Footwear
6. Grass

...

Annotation: Dog

System 2

1. Poodle
2. Footwear
3. Grass
4. Horse
5. Dog
6. Cat

...

1. {Dog, Chihuahua, Poodle}
2. {Sheep}
3. {Footwear, Sandal, Flip-flop}
4. {Grass, Pasture}
5. {Horse, Pony}

...

1. {Dog, Chihuahua, Poodle}
2. {Footwear, Sandal, Flip-flop}
3. {Grass, Pasture}
4. {Horse, Pony}
5. {Cat, Kitty}

...

‘Granularity aware’ groupings

- Group semantically related concepts
- *Across* varying levels of granularity
- Used at **evaluation** time
  - Initial rankings are altered in a manner that gives different insights

Related Work

- Basic-level categories (Biederman 1995)
- Differences in abstraction level when labelling groups of images vs. describing individual images (Rorissa 2008)
- Classifiers decide the best level of abstraction (Deng et al 2012)
- Learn most ‘natural’ basic-level category from text corpora (Ordonez et al 2013)
- We focus on *evaluating* systems across *multiple* levels of granularity

Method Outline

Cat
Dog
Footwear
Grass
Poodle
Sandal
…

Semantic Grouping

{Cat}
{Dog, Poodle}
{Footwear, Sandal}
{Grass, Pasture}
{Horse, Pony}
…

A ***dog*** is running away with a ***sandal*** in its mouth.
A large white ***poodle*** is walking on the grass carrying a ***sandal***.
A large white ***poodle*** running along with a pink ***flip flop*** in its mouth.
A white ***dog*** carries a pink ***flip-flop*** through the green yard.
Large white ***poodle*** carrying a pink ***flip flop*** outside.

Human Descriptions to 'Gold Standard'

{Dog, Poodle}
{Footwear, Sandal}

1. Poodle
2. Footwear
3. Grass
4. Horse
5. Dog

…

Reranking Classifier Output

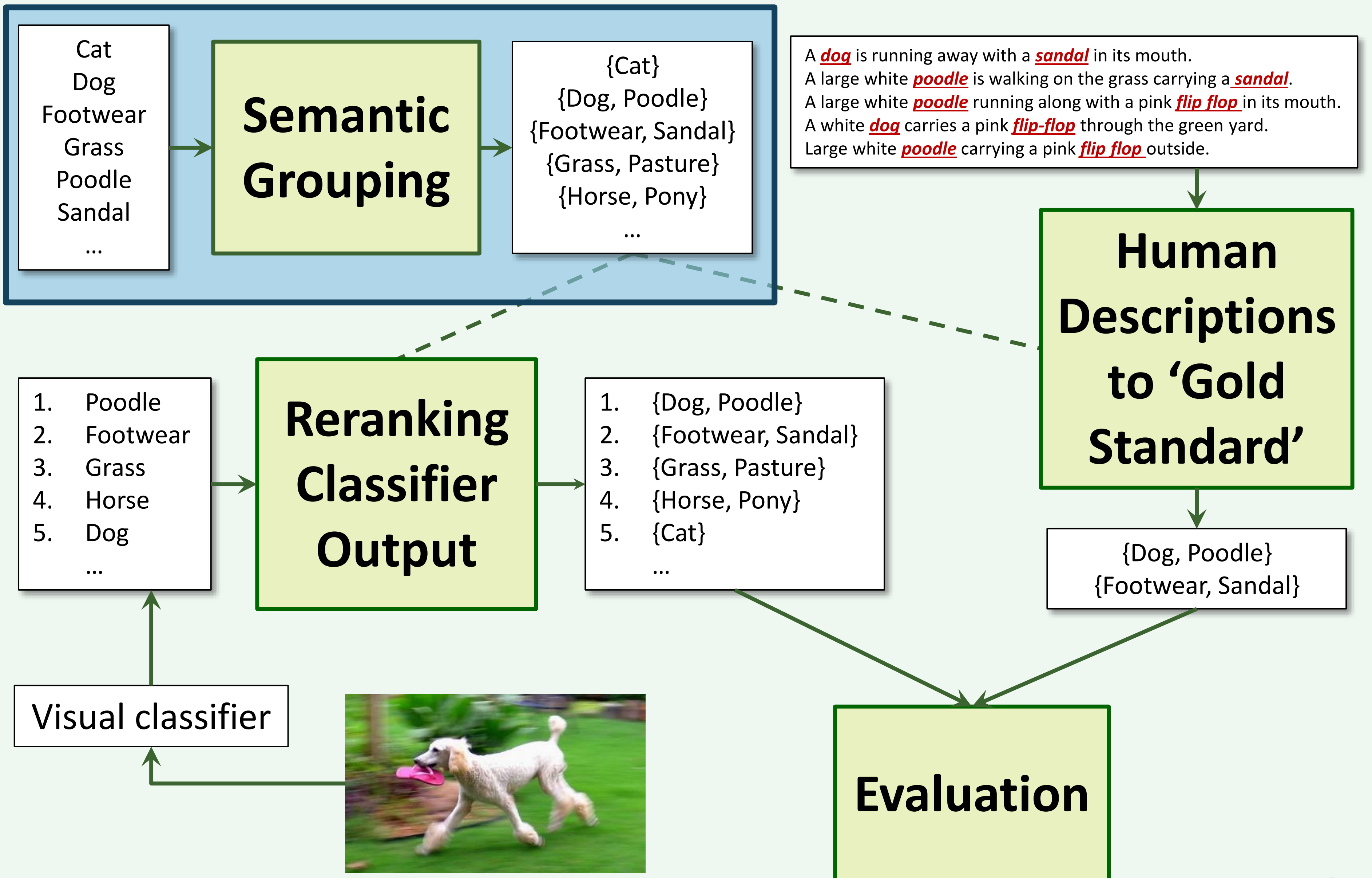
1. {Dog, Poodle}
2. {Footwear, Sandal}
3. {Grass, Pasture}
4. {Horse, Pony}
5. {Cat}

…

Visual classifier

Evaluation

Semantic Grouping



Semantic Grouping

Translation function from concept *v* to basic-level concepts (Ordonez et al. 2013)

ϕ	Concept	ψ
$\phi(being) = ln(235150)$	**being**	$\psi(being, poodle) = 3$
$\phi(animal) = ln(325501)$	**animal**	$\psi(animal, poodle) = 2$
$\phi(dog) = ln(406321)$	**dog**	$\psi(dog, poodle) = 1$
$\phi(poodle) = ln(6448)$	**poodle**	$\psi(poodle, poodle) = 0$

$$\tau(v, \lambda) = \arg\max_{w \in \Pi(v)} \left[\lambda\, \phi(w) - (1 - \lambda)\, \psi(w, v) \right]$$

hypernyms of *v*

'naturalness' (YFCC100M dataset)

semantic distance (WordNet)

Semantic Grouping

Group all concepts *v* translating to the same hypernym *w*

dog

$$\tau(\text{chihuahua}, \lambda) = \text{dog}$$
$$\tau(\text{dog}, \lambda) = \text{dog}$$
$$\tau(\text{poodle}, \lambda) = \text{dog}$$
$$\tau(\text{terrier}, \lambda) = \text{dog}$$

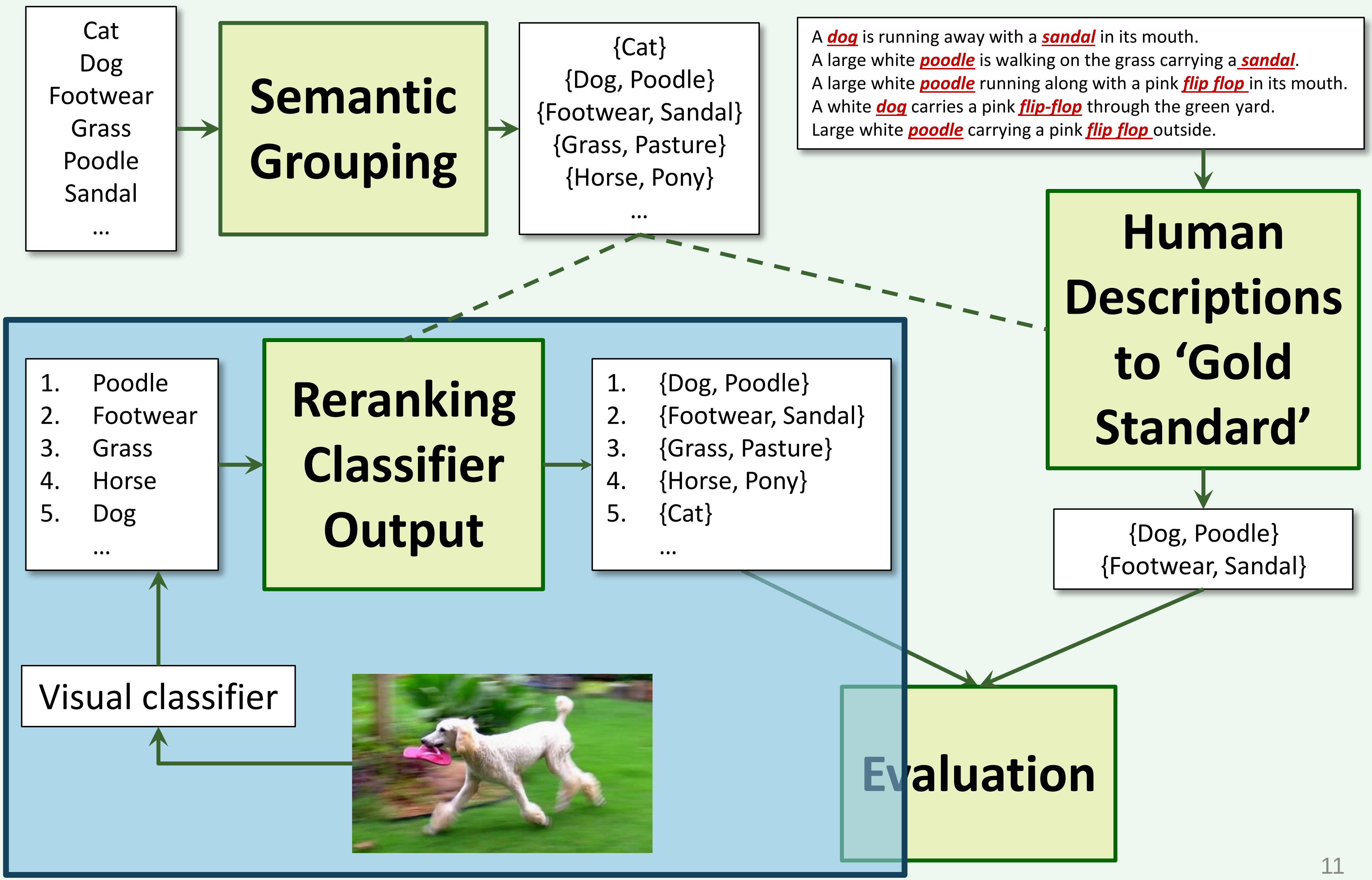
footwear

$$\tau(\text{footwear}, \lambda) = \text{footwear}$$
$$\tau(\text{sandal}, \lambda) = \text{footwear}$$

animal

$$\tau(\text{animal}, \lambda) = \text{animal}$$
$$\tau(\text{creature}, \lambda) = \text{animal}$$
$$\tau(\text{mussel}, \lambda) = \text{animal}$$

Reranking Classifier Output



Reranking Classifier Output

$\boldsymbol{\lambda = 0.5}$

Visual classifiers

1. poodle (0.92)
2. footwear (0.81)
3. grass (0.80)
4. horse (0.75)
5. dog (0.74)

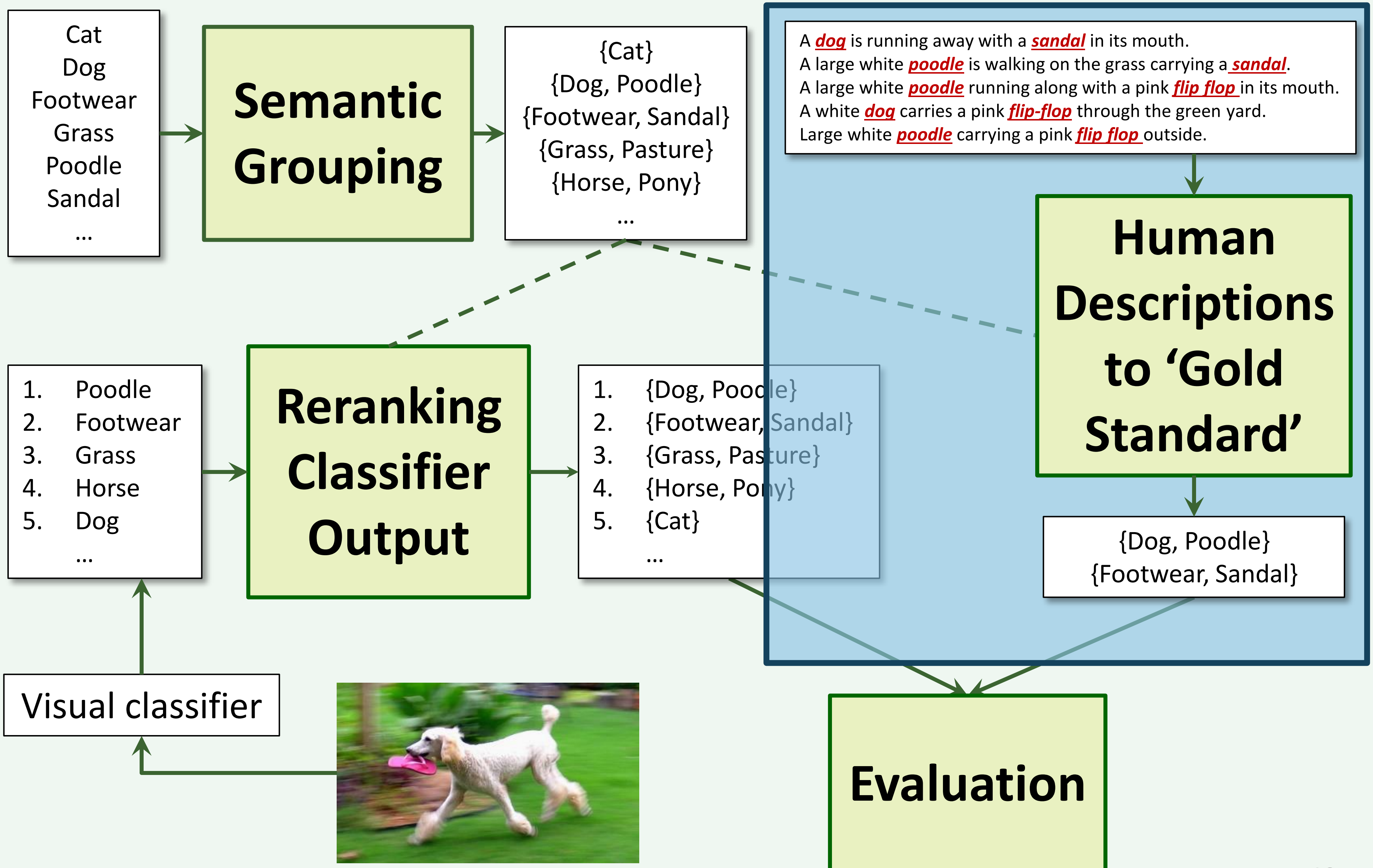
…

dog: {chihuahua, dog, poodle}
footwear: {footwear, sandal}
frisbee: {frisbee}
horse: {cob, horse, pony}
pasture: {pasture, grass}
animal: {animal, creature, mussel}
…

1. ***dog***: {chihuahua, dog, poodle} (0.92)
2. ***footwear***: {footwear, sandal} (0.81)
3. ***pasture***: {pasture, grass} (0.80)
4. ***horse***: {cob, horse, pony} (0.75)
5. ***frisbee***: {frisbee} (0.69)

…

Human Descriptions to 'Gold Standard'



Human Descriptions to 'Gold Standard'

- Flickr8k Dataset (Hodosh et al. 2013)
  - 8,091 images, 5 descriptions each

A dog is running away with a sandal in its mouth.
A large white poodle is walking on the grass carrying a sandal.
A large white poodle running along with a pink flip flop in its mouth.
A white dog carries a pink flip-flop through the green yard.
Large white poodle carrying a pink flip flop outside.

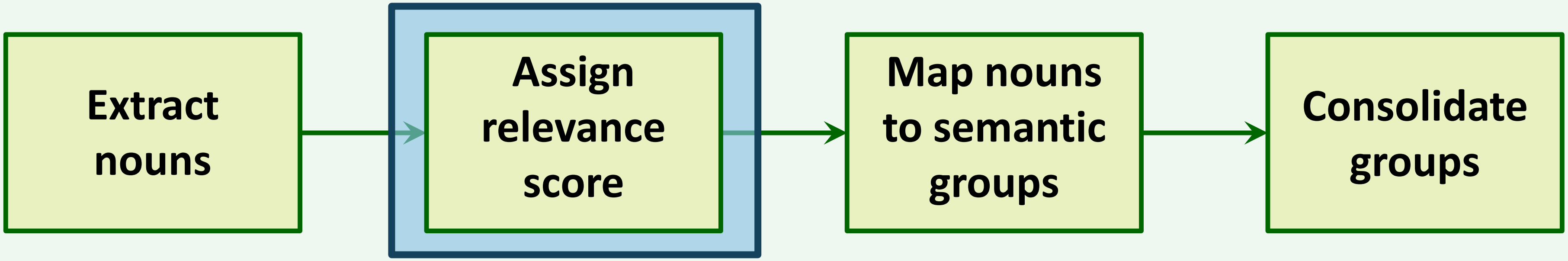
Human Descriptions to 'Gold Standard'

Extract nouns → **Assign relevance score** → **Map nouns to semantic groups** → **Consolidate groups**

A ***dog*** is running away with a ***sandal*** in its ***mouth***.
A large white ***poodle*** is walking on the ***grass*** carrying a ***sandal***.
A large white ***poodle*** running along with a pink ***flip flop*** in its ***mouth***.
A white ***dog*** carries a pink ***flip-flop*** through the green ***yard***.
Large white ***poodle*** carrying a pink ***flip flop*** outside.

dog **poodle** **sandal** **flip-flop** **grass** **yard** **mouth**

Human Descriptions to 'Gold Standard'



A ***dog*** is running away with a ***sandal*** in its ***mouth***.
A large white ***poodle*** is walking on the ***grass*** carrying a ***sandal***.
A large white ***poodle*** running along with a pink ***flip flop*** in its ***mouth***.
A white ***dog*** carries a pink ***flip-flop*** through the green ***yard***.
Large white ***poodle*** carrying a pink ***flip flop*** outside.

dog	poodle	sandal	flip-flop	grass	yard	mouth
2	3	2	3	2	1	2

Human Descriptions to 'Gold Standard'

Extract nouns → Assign relevance score → **Map nouns to semantic groups** → Consolidate groups

A ***dog*** is running away with a ***sandal*** in its ***mouth***.
A large white ***poodle*** is walking on the ***grass*** carrying a ***sandal***.
A large white ***poodle*** running along with a pink ***flip flop*** in its ***mouth***.
A white ***dog*** carries a pink ***flip-flop*** through the green ***yard***.
Large white ***poodle*** carrying a pink ***flip flop*** outside.

dog	poodle	sandal	flip-flop	grass	yard	mouth
2	3	2	3	2	1	2
dog	*dog*	*footwear*	*footwear*	*pasture*	*yard*	*mouth*

Human Descriptions to 'Gold Standard'

Extract nouns → **Assign relevance score** → **Map nouns to semantic groups** → **Consolidate groups**

A ***dog*** is running away with a ***sandal*** in its ***mouth***.
A large white ***poodle*** is walking on the ***grass*** carrying a ***sandal***.
A large white ***poodle*** running along with a pink ***flip flop*** in its ***mouth***.
A white ***dog*** carries a pink ***flip-flop*** through the green ***yard***.
Large white ***poodle*** carrying a pink ***flip flop*** outside.

dog	poodle	sandal	flip-flop	grass	yard	mouth
2	3	2	3	2	1	2

dog	*footwear*	*pasture*	*yard*	*mouth*
3	*3*	*2*	*1*	*2*

Evaluation

Cat
Dog
Footwear
Grass
Poodle
Sandal
…

Semantic Grouping

{Cat}
{Dog, Poodle}
{Footwear, Sandal}
{Grass, Pasture}
{Horse, Pony}
…

A ***dog*** is running away with a ***sandal*** in its mouth.
A large white ***poodle*** is walking on the grass carrying a ***sandal***.
A large white ***poodle*** running along with a pink ***flip flop*** in its mouth.
A white ***dog*** carries a pink ***flip-flop*** through the green yard.
Large white ***poodle*** carrying a pink ***flip flop*** outside.

Human Descriptions to 'Gold Standard'

{Dog, Poodle}
{Footwear, Sandal}

1. Poodle
2. Footwear
3. Grass
4. Horse
5. Dog

…

Reranking Classifier Output

1. {Dog, Poodle}
2. {Footwear, Sandal}
3. {Grass, Pasture}
4. {Horse, Pony}
5. {Cat}

…

Visual classifier

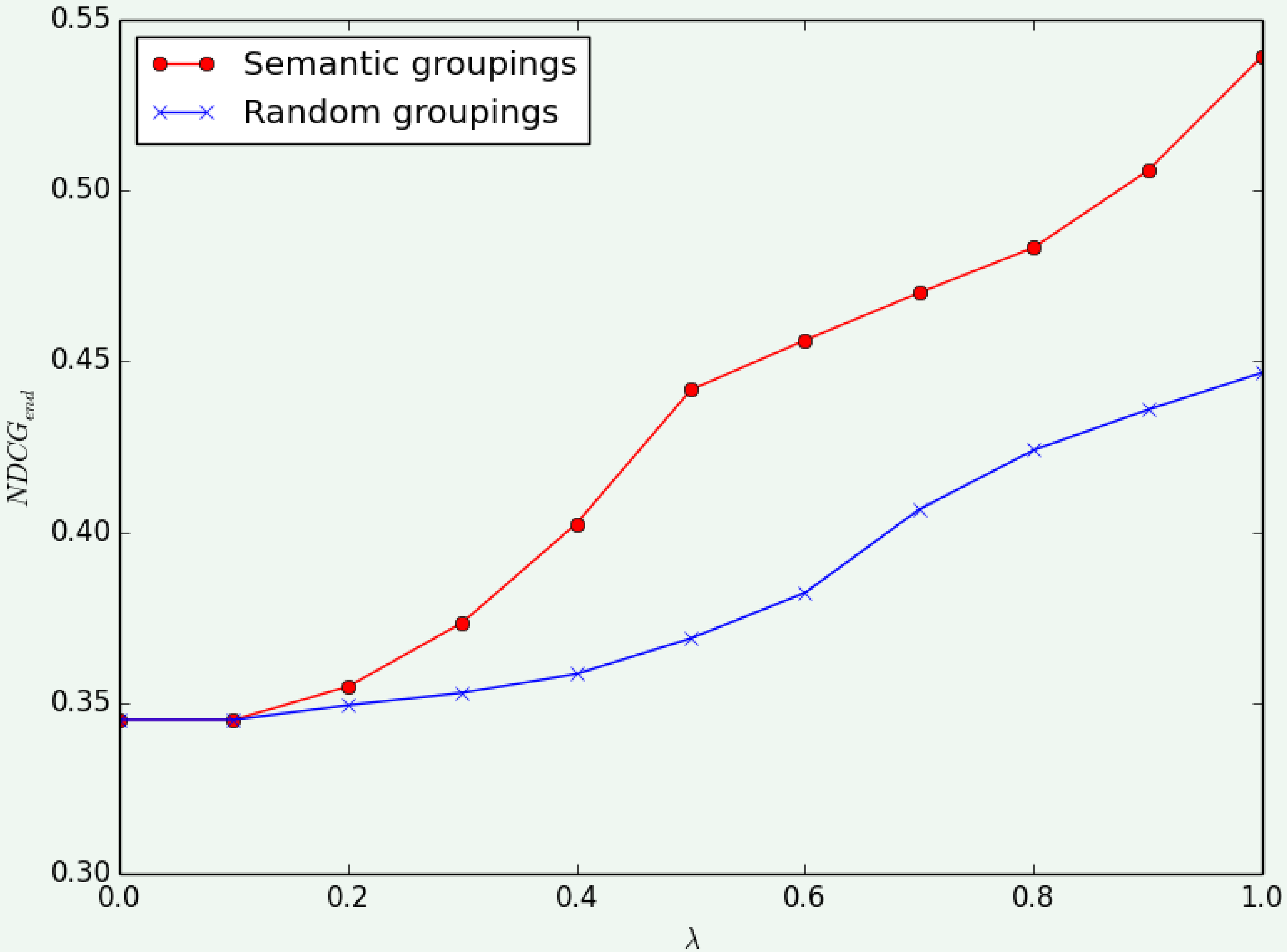
Evaluation

Evaluation Measure

- Normalized Discounted Cumulative Gain (NCDG)
  - Favours most relevant items first
  - Does not penalise irrelevant items
  - Normalised: between 0.0 to 1.0
    - Comparable across rankings with different no. of items

- Baseline: Concepts are grouped randomly

Results



Semantic groupings
Random groupings
0.55
0.50
0.45
0.40
0.35
0.30
0.0
0.2
0.4
0.6
0.8
1.0
$NDCG_{end}$
λ

Results

dog (5)
road (2)
sidewalk* (1)
street (1)

$\boldsymbol{\lambda = 0.0}$

beagle
boston_terrier
corgi
basset
hound
spaniel
border_collie
terrier
dachshund
pup
st_benard
bulldog
springer_spaniel
leash
kitten
pet
<u>dog</u>
sheepdog
penguin
<u>sidewalk</u>

sidewalk: **{pavement, sidewalk}**

Results

dog (5)
road (2)
sidewalk* (1)
street (1)

sidewalk: **{pavement, sidewalk}**

$\lambda = 0.0$	$\lambda = 0.3$
beagle	beagle
boston_terrier	boston_terrier
corgi	***dog***
basset	basset
hound	hound
spaniel	spaniel
border_collie	border_collie
terrier	terrier
dachshund	dachshund
pup	pup
st_benard	bulldog
bulldog	springer_spaniel
springer_spaniel	leash
leash	kitten
kitten	pet
pet	penguin
dog	***sidewalk***
sheepdog	doberman
penguin	collie
sidewalk	cat

Results

dog (5)
road (2)
sidewalk* (1)
street (1)

$\lambda = 0.3$	$\lambda = 0.5$
beagle	***dog***
boston_terrier	animal
dog	leash
basset	kitten
hound	pet
spaniel	penguin
border_collie	***sidewalk***
terrier	cat
dachshund	artifact
pup	person
bulldog	student
springer_spaniel	goat
leash	livestock
kitten	rabbit
pet	duck
penguin	baseball
sidewalk	chair
doberman	child
collie	frisbee
cat	spectator

sidewalk: **{pavement, sidewalk}**

Results

dog (5)
road (2)
sidewalk* (1)
street (1)

$\boldsymbol{\lambda = 0.5}$

<u>dog</u>
animal
leash
kitten
pet
penguin
<u>sidewalk</u>
cat
artifact
person
student
goat
livestock
rabbit
duck
baseball
chair
child
frisbee
spectator

$\boldsymbol{\lambda = 0.8}$

<u>dog</u>
animal
leash
being
bird
<u>sidewalk</u>
cat
artifact
student
baseball
chair
child
frisbee
ball
slope
equipment
fabric
rug
seat
support

sidewalk: **{pavement, sidewalk}**

Results

boat (4)
graham (1)
raceway* (1)
vessel* (1)

raceway: **{race, raceway}**
vessel: **{vessel, watercraft}**

$\lambda = 0.0$	$\lambda = 0.3$
motorboat	***boat***
speedboat	lifeguard
lifeguard	lifeboat
lifeboat	car
racer	***vessel***
vessel	sidecar
car	kayak
barge	tugboat
sidecar	paddle
kayak	dinghy
boat	motor
tugboat	ferry
paddle	glass
dinghy	preserver
motor	***raceway***
ferry	oar
bumper	airplane
preserver	vehicle
raceway	raft
oar	hatchback

Results

boat (4)
graham (1)
raceway* (1)
vessel* (1)

raceway: **{race, raceway}**
vessel: **{vessel, watercraft}**

$\lambda = 0.3$	$\lambda = 0.5$
boat	***boat***
lifeguard	lifeguard
lifeboat	car
car	***vessel***
vessel	sidecar
sidecar	kayak
kayak	tugboat
tugboat	paddle
paddle	motor
dinghy	glass
motor	equipment
ferry	***raceway***
glass	oar
preserver	airplane
raceway	vehicle
oar	raft
airplane	hatchback
vehicle	device
raft	screen
hatchback	field

Results

***boat* (4)**
***graham* (1)**
***raceway* (1)**
<u>*vehicle*</u> *(1)*

vehicle: **{buggy, bulldozer, camper, carriage, cart, …, sailboat, …, vehicle, vessel, wagon, watercraft, wheelbarrow, yacht}**

$\lambda = 0.5$

<u>boat</u>
lifeguard
car
<u>vessel</u>
sidecar
kayak
tugboat
paddle
motor
glass
equipment
<u>raceway</u>
oar
airplane
vehicle
raft
hatchback
device
screen
field

$\lambda = 0.8$

<u>boat</u>
being
car
<u>vehicle</u>
artifact
tugboat
paddle
motor
glass
equipment
<u>raceway</u>
airplane
raft
structure
device
screen
field
wrapping
brush
pop

Discussion

- We proposed grouping semantically related concepts *across* different levels of granularity
- Semantic groups are used to alter visual classifier rankings, provides different insights
- Human-centric evaluation with descriptions
- Trade-off between flexibility vs. informativeness
- Future work:
  - Different methods of grouping concepts
  - Incorporate visual classifiers for grouping & reranking

josiahwang.com

A Poodle or a Dog ?

Evaluating Automatic Image Annotation Using Human Descriptions at Different Levels of Granularity

Josiah Wang Fei Yan Ahmet Aker Robert Gaizauskas

Department of Computer Science
The University Of Sheffield.

Centre for Vision, Speech and Signal Processing
UNIVERSITY OF SURREY

Work presented as part of the Visual Sense (ViSen) project, funded by the chist-era D2K programme