



# Situation Recognition: Visual Semantic Role Labelling for Image Understanding

Mark Yatskar, Luke Zettlemoyer, Ali Farhadi

Experiment presentation by Nayan Singhal



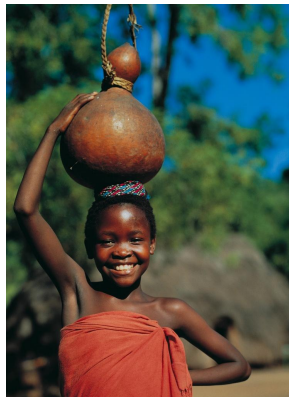
# Motivation

- Human understanding of image
- Verbs in English language.



# Approach

- CRF with CNN



1024



CARRYING	
AGENT	WOMAN
ITEM	JAR
AGENTPART	HEAD
PLACE	OUTDOOR

- Log linear Loss

$$\sum_{i \in D} \log \left( 1 - \prod_{S \in A_i} (1 - p(S|i; \theta)) \right)$$



# How object plays role in image understanding?



neighboring images

jumping				
agent	source	obstacle	destination	place
man	cliff	∅	cliff	outdoors

0.89821



# Remove Cliff



Removing Cliff



neighboring images

falling			
agent	source	goal	place
man	sky	land	outdoors

0.45862



# Remove person



Removing Man



neighboring images

cresting	
agent	place
cloud	outdoors

0.60973



# Remove Sky



Removing Sky



neighboring images

jumping				
agent	source	obstacle	destination	place
man	cliff	∅	cliff	outdoors

0.64695





# Image (2)



© North News & Pictures Ltd



neighboring images

0.85718

falling			
agent	source	goal	place
person	horse	land	outdoors



# Remove Person



© North News & Pictures Ltd

Removing man



neighboring images

	competing	
agent	competition	place
jockey	horse race	path

0.47155



# Remove Background



Removing Sky and Man



neighboring images

rearing	
agent	place
horse	grass

0.83053



# Conclusion

Each object plays a significant role in image understanding.



# Experiment

- 1) Analyzing Failure Cases
- 2) Different moods of faces



# Expt 1: Analyzing Failure Cases



# Object Recognition (1)



Downloaded from [www.istockphoto.com](http://www.istockphoto.com)

clawing		
agent	victim	place
mother	blackboard	∅

0.93069

clinging		
agent	clungto	place
cat	tree	outdoors

0.05893

rubbing			
agent	item	agentpart	place
cat	leg	wood	outside

0.00344

scratching			
agent	object	tool	place
cat	tree	claw	inside

0.00292

Insitu Result



# Object Recognition (2)



shutterstock

www.shutterstock.com · 48585736

leaping				
agent	source	obstacle	destination	place
male child	∅	∅	∅	∅

0.64129

dancing	
agent	place
couple	∅

0.07168

twisting		
agent	agentpart	place
woman	body	∅

0.06156

shouting		
agent	addressee	place
woman	∅	∅

0.05621

Insitu Result





# Object Recognition (3)



aiming			
agent	item	target	place
woman	firearm	∅	room

0.97364

shooting			
agent	projectile	firearm	target
man	bullet	rifle	∅

0.01296

detaining		
agent	victim	place
soldier	male child	outdoors

0.00293

guarding			
agent	item	weapon	place
soldier	∅	gun	outdoors

0.00220

Insitu Result



# Object Recognition (4)



drenching				
agent	drencheditem	liquid	tool	place
man	woman	water	∅	outdoors

0.50556

immersing			
agent	item	liquid	place
woman	∅	water	pool

0.39952

sliding				
agent	slider	surface	destination	place
woman	∅	slide	body of water	pool

0.02837

Imsitu Result



# Object Recognition (5)



clawing		
agent	victim	place
mother	blackboard	∅

0.52204

nuzzling		
agent	item	place
cow	cow	field

0.10331

chewing		
agent	item	place
horse	wood	outdoors

0.05848

Insitu Result



# Object Recognition (6)



installing				
agent	component	destination	tool	place
man	board	wall	screwdriver	room

0.35408

fastening					
agent	item	destination	connector	tool	place
man	∅	∅	∅	hand	inside

0.05477

installing			
agent	component	tool	place
woman	board	screwdriver	inside

0.04931

Imsitu Result



Why is it happening?

Are these images difficult to categorize?



Let's analyze these with ImageNet



# Object Recognition (1)



Downloaded from [www.istockphoto.com](http://www.istockphoto.com)

clawing		
agent	victim	place
mother	blackboard	∅

0.93069

clinging		
agent	clungto	place
cat	tree	outdoors

0.05893

rubbing			
agent	item	agentpart	place
cat	leg	wood	outside

0.00344

scratching			
agent	object	tool	place
cat	tree	claw	inside

0.00292

Insitu Result



Downloaded from [gettyimages.com](https://www.gettyimages.com)

- |                 |                 |
|-----------------|-----------------|
| cute 0.994...   | fur 0.994...    |
| stripe 0.989... | cat 0.988...    |
| animal 0.988... | nature 0.985... |
| eye 0.974...    | little 0.974... |
| kitten 0.966... | pet 0.961...    |
| sit 0.958...    | young 0.955...  |

ImageNet classification





# Object Recognition (2)



shutterstock

www.shutterstock.com · 48585736

leaping				
agent	source	obstacle	destination	place
male child	∅	∅	∅	∅

0.64129

dancing	
agent	place
couple	∅

0.07168

twisting		
agent	agentpart	place
woman	body	∅

0.06156

shouting		
agent	addressee	place
woman	∅	∅

0.05621

Insitu Result



# Object Recognition (2)



www.shutterstock.com · 48585736

- woman 0.992...
- exercise 0.985...
- sport 0.980...
- action 0.976...
- young 0.972...
- aerobics 0.964...
- strength 0.961...
- girl 0.956...
- body 0.956...
- beautiful 0.949...
- gymnastics 0.945...
- fitness 0.940...

ImageNet classification



# Object Recognition (3)



aiming			
agent	item	target	place
woman	firearm	∅	room

0.97364

shooting			
agent	projectile	firearm	target
man	bullet	rifle	∅

0.01296

detaining		
agent	victim	place
soldier	male child	outdoors

0.00293

guarding			
agent	item	weapon	place
soldier	∅	gun	outdoors

0.00220

Insitu Result



# Object Recognition (3)



- camouflage 0.999...
- army 0.998...
- military 0.998...
- war 0.997...
- gun 0.993...
- weapon 0.992...
- battle 0.990...
- rifle 0.988...
- uniform 0.986...
- soldier 0.986...
- combat 0.977...
- protection 0.970...
- khaki 0.967...
- man 0.956...

ImageNet classification



# Object Recognition (4)



drenching				
agent	drencheditem	liquid	tool	place
man	woman	water	∅	outdoors

0.50556

immersing			
agent	item	liquid	place
woman	∅	water	pool

0.39952

sliding				
agent	slider	surface	destination	place
woman	∅	slide	body of water	pool

0.02837

Imsitu Result



# Object Recognition (4)



- water 0.985...
- wet 0.984...
- leisure 0.978...
- relaxation 0.974...
- swimmer 0.968...
- recreation 0.963...
- summer 0.949...
- woman 0.948...
- people 0.935...
- vacation 0.926...
- resort 0.917...

ImageNet classification



# Object Recognition (5)



clawing		
agent	victim	place
mother	blackboard	∅

0.52204

nuzzling		
agent	item	place
cow	cow	field

0.10331

chewing		
agent	item	place
horse	wood	outdoors

0.05848

Insitu Result



# Object Recognition (5)



- mammal 0.998...
- animal 0.997...
- farm 0.993...
- grass 0.991...
- cow 0.991...
- livestock 0.989...
- nature 0.986...
- cattle 0.985...
- hayfield 0.984...
- agriculture 0.973...
- field 0.972...
- milk 0.971...
- no person 0.965...
- rural 0.965...

ImageNet classification





# Object Recognition (6)



installing				
agent	component	destination	tool	place
man	board	wall	screwdriver	room

0.35408

fastening					
agent	item	destination	connector	tool	place
man	∅	∅	∅	hand	inside

0.05477

installing			
agent	component	tool	place
woman	board	screwdriver	inside

0.04931

Imsitu Result



# Object Recognition (6)

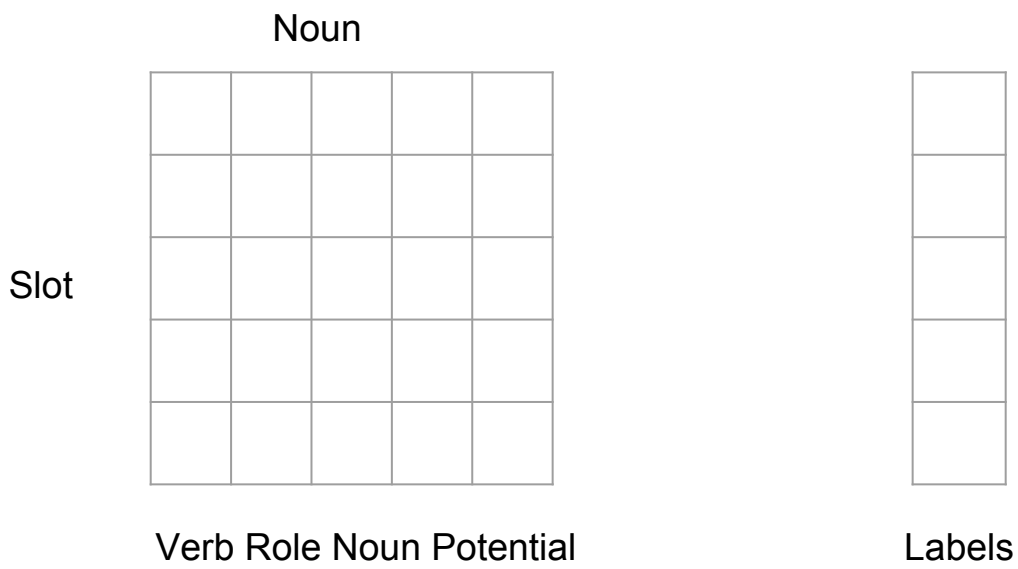


- one 0.987...
- people 0.987...
- indoors 0.983...
- woman 0.982...
- adult 0.974...
- furniture 0.963...
- carpentry 0.936...
- room 0.933...
- wear 0.925...
- wood 0.912...

ImageNet classification



# Object Recognition (1)



Post Processing: ~~**A** (Verb Role Noun Potential) + **B** (Labels)~~



# Object Recognition (1)



Downloaded from [www.istockphoto.com](http://www.istockphoto.com)

clawing		
agent	victim	place
mother	blackboard	∅

0.93069

clinging		
agent	clungto	place
cat	tree	outdoors

0.05893

rubbing			
agent	item	agentpart	place
cat	leg	wood	outside

0.00344

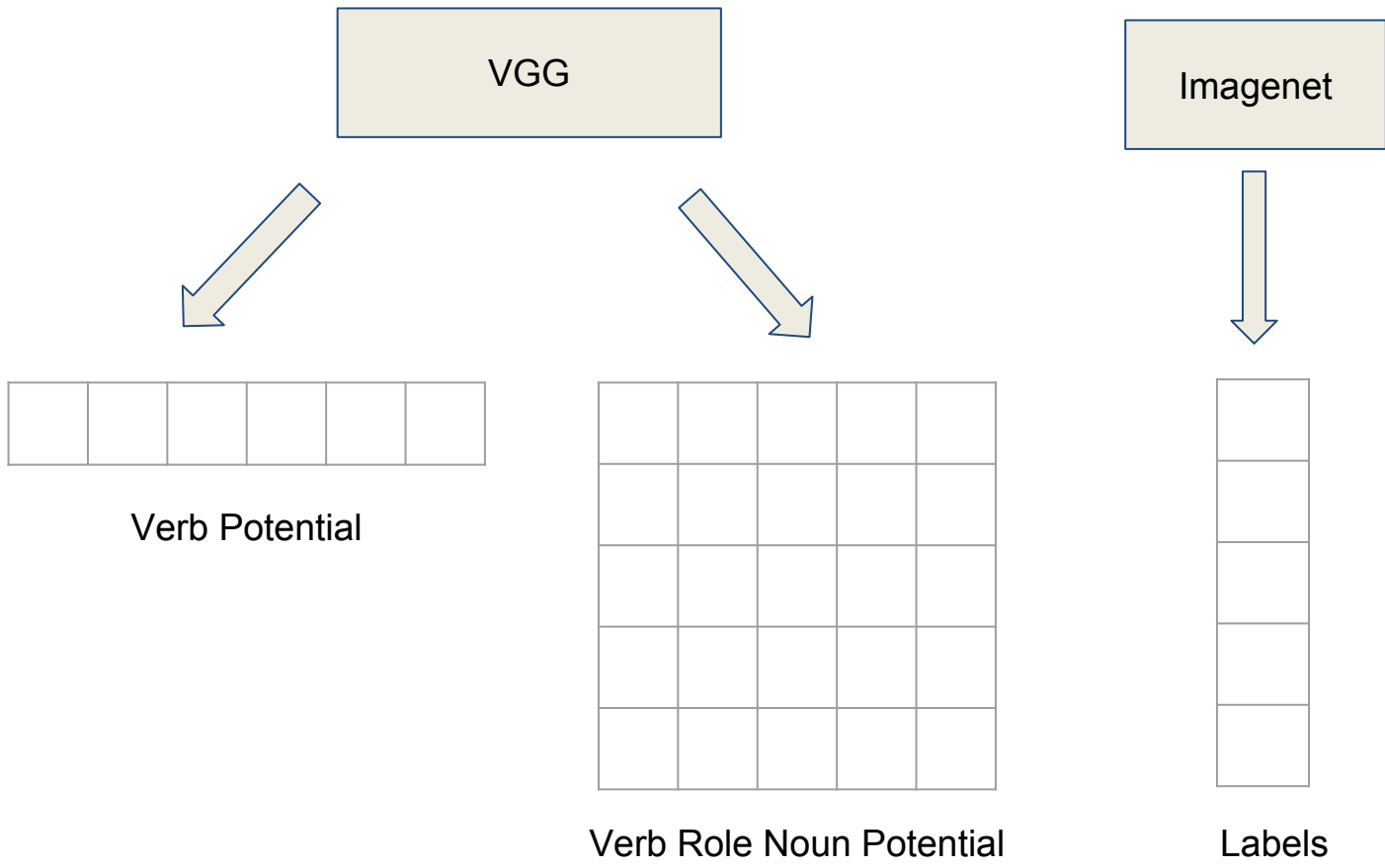
scratching			
agent	object	tool	place
cat	tree	claw	inside

0.00292

Insitu Result



# Object Recognition (1)



**Preprocessing**



# Future Work

- Add labels in preprocessing.



# Exp 2: Different moods

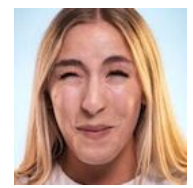
- Laughing
- Smiling
- Frowning
- Grimacing
- Winking
- Squinting
- Shouting
- Puckering



Winking



Squinting



Puckering



Smiling



Laughing



Frowning



# Success Case



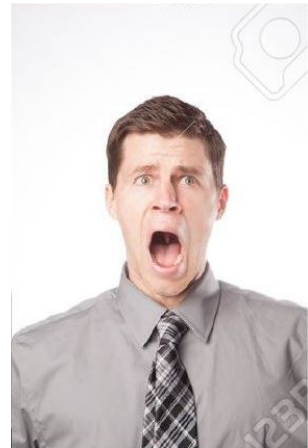
Smiling	
Agent	place
man	-

0.35967



Laughing	
Agent	place
man	-

0.35777



Shouting	
Agent	place
man	-

0.37531





# Success Case



Frowning	
Agent	place
man	-

0.24378



Grimacing	
Agent	place
man	-

0.21052



# Failure Case



Winking	
Agent	place
man	-

0.20954

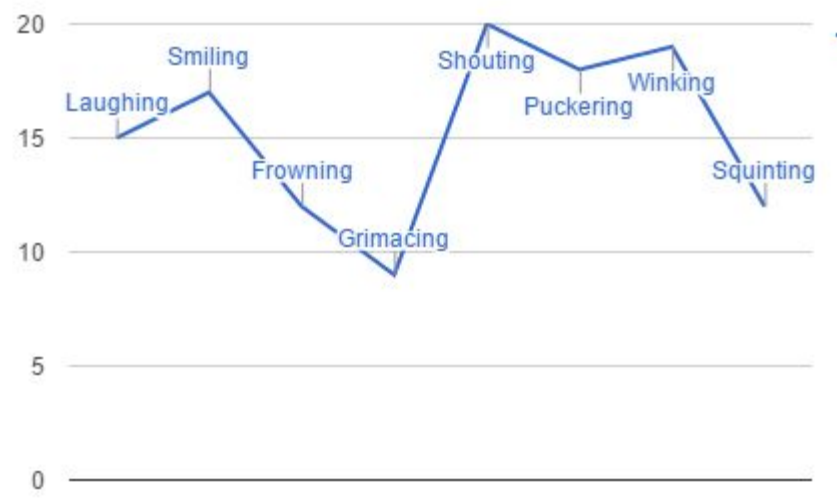


Puckering	
Agent	place
woman	-

0.21052



Test Images (25)



- Conclusion: Detect different moods of faces with slight variation.



# Some Interest Categorization



laughing	
agent	place
man	outside

0.53253



# Some Interest Categorization



Camouflaging	
Agent	owl
Hiding Item	tree
Place	outdoors



Camouflaging	
Agent	frog
Hiding Item	pebble
Place	-



neighboring images

vaulting					
agent	start	obstacle	end	tool	place
man	land	pommel horse	land	person	outdoors

0.20864



neighboring images

guarding			
agent	item	weapon	place
dog	sheep	∅	field

0.56284



**Thank You**





# No Agent(2)



Watering	
Agent	Person
Tool	Bucket
Place	garden



Shredding	
Agent	Person
Tool	Shreder
Item	paper
Place	-