Two professors converse in front of a blackboard.
Two professors stand in front of a blackboard.
Two professors converse in front of a blackboard.
Is photorealism necessary?
Jenny

Mike
How do we generate scenes?

Create a children's illustration!

Please help us create an illustration for a children's story book by creating a realistic scene from the clipart below. Use your imagination! Clipart may be added by dragging the clipart onto the scene, and removed by dragging it off. The clipart may be resized or flipped, and each clipart may only be added once. Please use at least 6 pieces of clipart in each scene. You will be asked to complete 3 different scenes. Press "Next" when finished with the current scene and "Done" when all are finished. Thanks!

<table>
<thead>
<tr>
<th>Size</th>
<th>Flip</th>
<th>Clipart</th>
</tr>
</thead>
<tbody>
<tr>
<td>😊😊😊</td>
<td>🧐🧐🧐</td>
<td><img src="image-url" alt="Clipart" /></td>
</tr>
</tbody>
</table>

Scene 1/3
Jenny loves to play soccer but she is worried that Mike will kick the ball too hard.

Mike and Jenny play outside in the sandbox. Mike is afraid of an owl that is in the tree.
Previous work

Sentence generation

Farhadi et al., Every picture tells a story: Generating sentences from images. ECCV, 2010.
Ordonez et al., Im2text: Describing images using 1 million captioned photographs. NIPS, 2011.
Yang et al., Corpus-guided sentence generation of natural images. EMNLP, 2011.
Kulkarni et al., Baby talk: Understanding and generating simple image descriptions. CVPR, 2011.
Kuznetsova et al., Collective Generation of Natural Image Descriptions. ACL, 2012.
Gupta et al., Choosing Linguistics over Vision to Describe Images. AAAI, 2012.

Nouns

Spain and Perona, Measuring and predicting object importance. IJCV 2011.
Hwang and Grauman, Learning the relative importance of objects... IJCV, 2011.

Adjectives, prepositions

Gupta and Davis, Beyond nouns ..., ECCV, 2008.
Farhadi et al., Describing objects by their attributes. CVPR, 2009.
Berg et al., Automatic attribute discovery and characterization from noisy web data. ECCV 2010.

Verbs

Yao and Fei-Fei, Modeling mutual context ... in human-object interaction activities. CVPR 2010.
Sadeghi and Farhadi, Recognition using visual phrases. CVPR 2011.
“Jenny just threw the beach ball angrily at Mike while the dog watches them both.”
Mike fights off a bear by giving him a hotdog while jenny runs away.
Jenny and Mike are both playing dangerously in the park.
Semantic importance of visual features

1,000 classes of semantically similar scenes:

Class 1

Class 2

Class 1,000

1,000 classes x 10 scenes per class = 10,000 scenes
Visual features
Visual features

Cloud

Tree

Cat

Basketball

Person standing

Person sitting

Gaze

Smile
Which visual features are semantically meaningful?

- Absolute position
- Absolute depth
- Relative depth
- Expression
- Co-occurrence
- Attachment to hand/head
- Category vs. instance
- Relative position
- Occurrence
- Pose
- Gaze
Which words are visually meaningful?

Happy
Very
Distinguished
Today
Basketball
To
A
Help
Bike
Kicking
Cloud
Vision
Mutual information

\[ I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \]

- Visual features
- Semantic classes
- Words
Information shared between:

Visual features
&
Semantic classes
Object occurrence

High

Low

Mutual Information

Mutual Information
Person attributes

Person Attributes

- Angry
- Smile
- Sit
- Happy
- Surprise
- Sad
- Kick
- Run
- Jump
- Stand
- Cross
- Wave

Conditional MI

0 0.2 0.4 0.6 0.8
Relative spatial

Relative orientation is very informative.
Information shared between:

**Visual features**
&

**Words**
Most visually informative words
Least visually informative words

today
home
me
something
attention
using
isn’t
doing
went
give
behind
before
during
onto
through
how
since
why
finally
almost
Most informative of relative position

bear
away
ball
soccer

kicking
Mike
from
on
his
to

he
him
holding
sandbox
a
playing
Most informative of relative position
What did we learn?

- Occurrence of object instances provides significant semantic information
- Frequency of occurrence ≠ semantic importance
- Human expression and pose are important attributes
- Occurrence of objects = nouns, while relative position is more predictive of verbs, adverbs and prepositions
- Relative position is more important than absolute position
- Co-occurrence of the boy/girl and animals are important
- ...

Duh, we already know that...

...but I didn’t.
What did we learn?

New approach to learning “common sense” knowledge about our world.

Goes beyond “Jenny and Mike.”
Jenny loves to play soccer but she is worried that Mike will kick the ball too hard. Mike and Jenny play outside in the sandbox. Mike is afraid of an owl that is in the tree. A cat anxiously sits in the park and stares at a unattended hot dog that someone left on a yellow bench. Mike is playing with a soccer ball when he is almost struck by lightening. Jenny had a pie that she didn't want to share. That made Mike angry. A balloon flies past them. Mike and Jenny are playing catch with a football while a dog watches and a hot air balloon flies past them. Mike and Jenny are having a great time in the sunny park as she pitches a baseball to Mike who is waiting with his bat. Jenny is talking to an owl in the tree. The owl is actually a wizard that is disguised.

Don't wait!
Thanks!

Special thanks to Bryan Russell, Lucy Vanderwende, Michel Galley, Luke Zettlemoyer
Everyone freaked out when Mike started levitating.
After watching The Help, Mike was cautious with pies.
After watching The Help, Mike became cautious with pies. Having ruined a few shirts, Mike became cautious with pies.