

Corpus-Guided Sentence Generation of Natural Images

Yezhou Yang* Ching L. Teo* Hal Daume and Yiannis Aloimonos
University of Maryland Institute for Advanced Computer Studies



UMIACS

UNIVERSITY OF MARYLAND INSTITUTE FOR ADVANCED COMPUTER STUDIES

What happens when you see a Picture?

Visual Space

Perception



Grounding

Language Space

World Knowledge

nouns

verbs

adjectives

prepositions

adverbs

Production



Two cows in a field grazing near a gate.
The large cows hover over the young calf.
Three adult cows and one baby cow stand on the grass.
Three brown cows and a small calf in a field.
Three cows in a green pasture surrounding a baby cow.

Speech/Text Generation

What is a descriptive sentence for an image?

- 1) the important *objects* (Nouns) that participate in the image;
- 2) Some description of the *actions* (Verbs) associated with these objects;
- 3) The *scene* where this image was taken;
- 4) the *preposition* that relates the objects to the scene.

$$T = \{n, v, s, p\}$$

Challenges



(a)



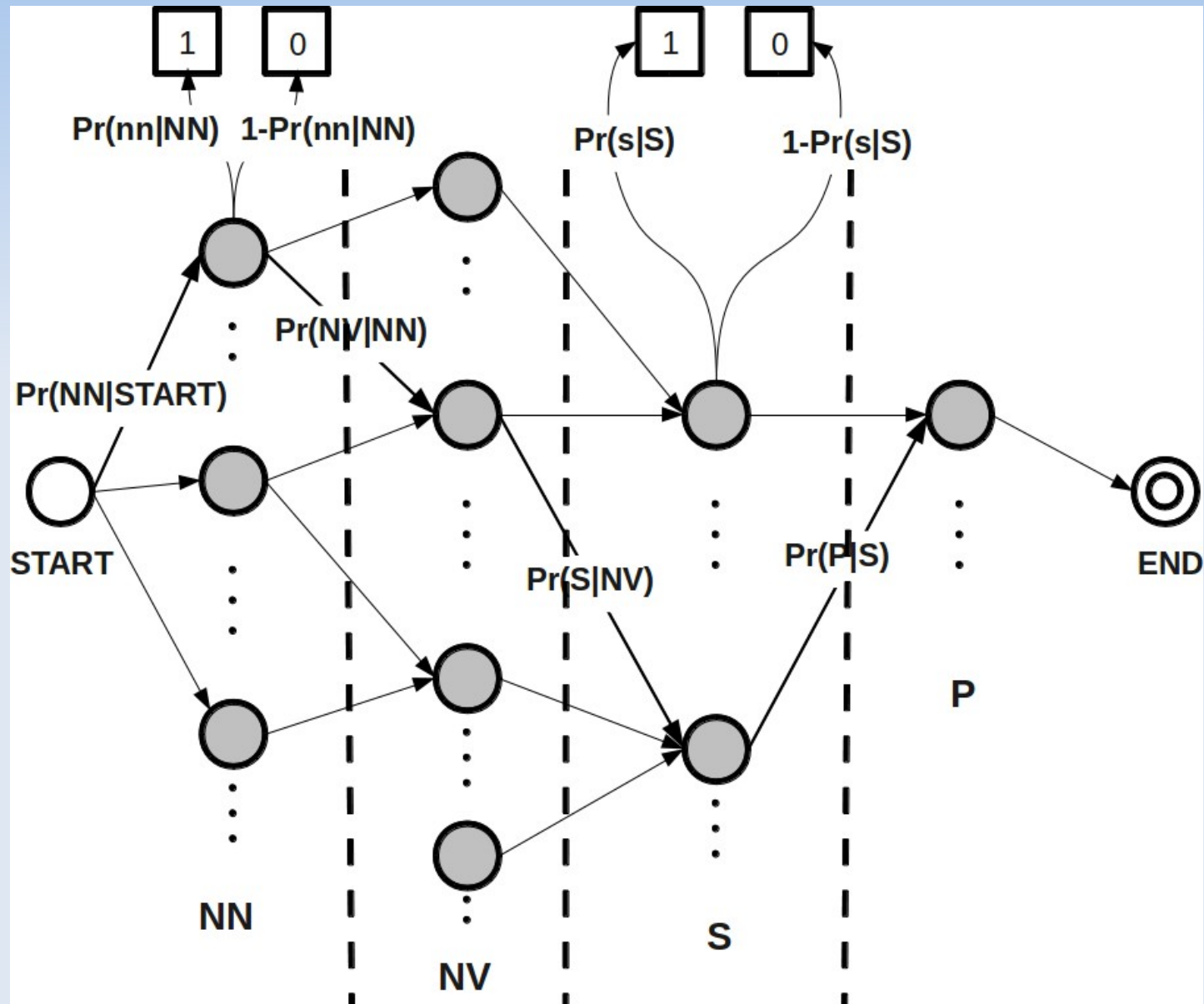
(b)

Overview of our approach

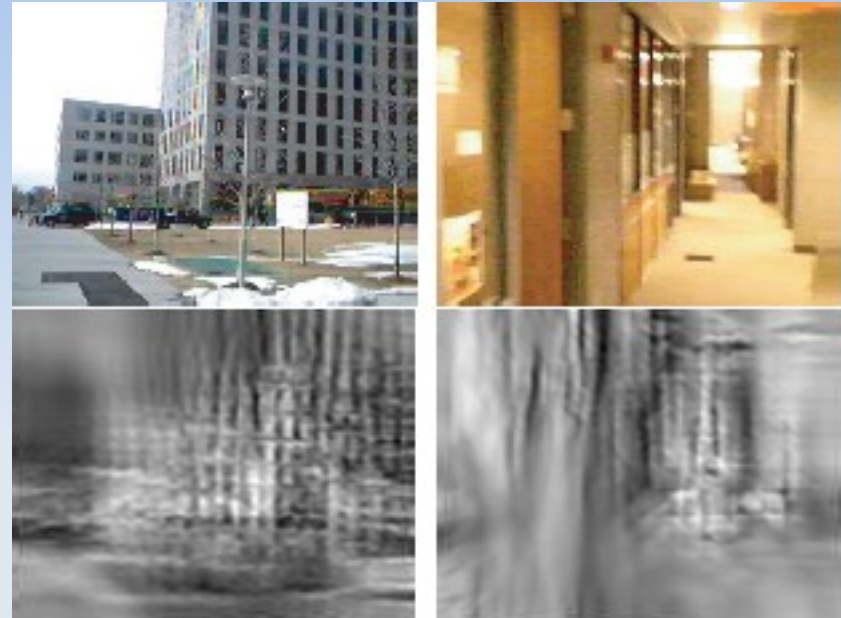
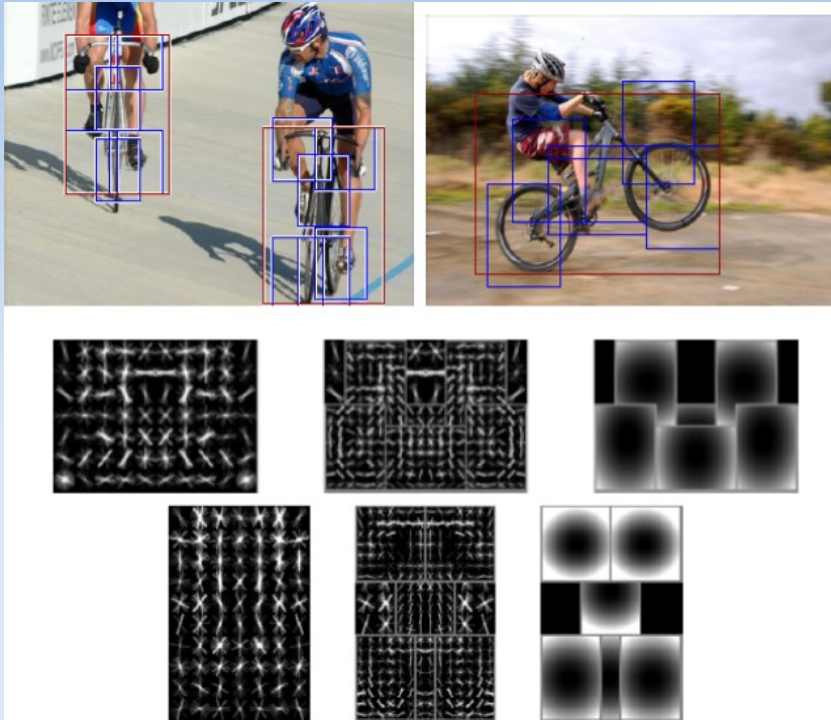


- a) Detect objects and scenes from input image;
- b) Estimate optimal sentence structure quadruplet T ;
- c) Generating a sentence from T ;

Determining T^* using HMM inference



Object and Scene Detections



Left: The part based object detector $\Pr(n|I)$;
Right: The GIST gradients based scene detector $\Pr(s|I)$;

UIUC PASCAL Sentence Dataset



The cow is grazing in a field.

An ox stands in a field

A yak with a long, camel colored coat standing in a field.

A young highlander cow stands in a pasture.

Closeup of a bull with hair covering its eyes



an Asian woman sitting in a chair on her balcony

A woman smiling.

Smiling Asian woman in floral dress.

The happy lady enjoys her surroundings.

The woman in the floral dress is posing among plants.



A dinner table set for three people.

A Thanksgiving meal with white daisies on a small table.

Dinner sitting on a table and ready to be served.

There is a turkey on the table along with other foods on plates.

The table is set for a turkey dinner and decorated-
with white daisies.

The set of objects, actions, scenes and prepositions

- Objects: 'aeroplane' 'bicycle' 'bird' 'boat' 'bottle' 'bus' 'car' 'cat' 'chair' 'cow' 'table' 'dog' 'horse', 'motorbike' 'person' 'pottedplant' 'sheep' 'sofa' 'train' 'tvmonitor'
- Actions: 'sit' 'stand' 'park' 'ride' 'hold' 'wear' 'pose' 'fly' 'lie' 'lay' 'smile' 'live' 'walk' 'graze' 'drive' 'play' 'eat' 'cover' 'train' 'close' ...
- Scenes: 'airport' 'field' 'highway' 'lake' 'room' 'sky' 'street' 'track'
- Preps: 'in' 'at' 'above' 'around' 'behind' 'below' 'beside' 'between' 'before' 'to' 'under' 'on'

Corpus-Guided Predictions

Predicting Verbs:

$$\text{Pr}(v|n1, n2) = \#(v,n1,n2)/\#(n1,n2);$$

Predicting Scenes:

$$\text{Pr}(s|n, v) = P(s|n)P(s|v);$$

$$P(s|n) = \#(s,n)/\#(n);$$

$$P(s|v) = \#(s,v)/\#(v);$$

Predicting Preps:

$$\text{Pr}(p|s) = \#(p,s)/\#(s);$$

Example:

*'the large brown **dog** **chases** a small young **cat** around the messy room, forcing the **cat** to **run** away towards its **owner**.'*

Sample Results



{aeroplane,fly,airport,at}
the aeroplane is flying at the airport.



{person,motorbike,ride,field,in}
the person is riding the motorbike in the field.



{person,bicycle,ride,street,on}
the person is **riding** the bicycle on the street.



{person,table,sit,room,in}
three people are sitting at the table in the room.

Turks evaluation

Rate these sentences. Are they good, average or bad?


Instructions:

IF YOU HAVE WORKED ON THIS HIT BEFORE, DO NOT ATTEMPT TO SUBMIT AGAIN! IT WILL BE AUTOMATICALLY REJECTED!

Return this HIT so that other workers have a chance.

We need a unique opinion per HIT for evaluation so a unique input is all that is needed.

Given an example image with 6 - 7 sentences that describe it, you are to rate the sentences by selecting the appropriate radio buttons:

Image	Sentences	Readability	Correctness
	1. Cows are grazing on the field	5 <input checked="" type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>	5 <input type="radio"/> 4 <input type="radio"/> 3 <input checked="" type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>
	2. Cows is graze on the field	5 <input type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input checked="" type="radio"/> 1 <input type="radio"/>	5 <input type="radio"/> 4 <input type="radio"/> 3 <input checked="" type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>
	3. A young calf is surrounded by cows	5 <input checked="" type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>	5 <input checked="" type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>
	4. Cows in the field	5 <input type="radio"/> 4 <input checked="" type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>	5 <input checked="" type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>
	5. Sheep and cows resting in field	5 <input type="radio"/> 4 <input checked="" type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>	5 <input type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input checked="" type="radio"/> 1 <input type="radio"/>
	6. Some animals outside	5 <input checked="" type="radio"/> 4 <input type="radio"/> 3 <input type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>	5 <input type="radio"/> 4 <input type="radio"/> 3 <input checked="" type="radio"/> 2 <input type="radio"/> 1 <input type="radio"/>

Sentences are rated on 2 scales: **Readability** and **Correctness**

- *Grammatically sound* sentences have better *readability* ratings. E.g. 'Cows are grazing on the field' is better than 'Cows is graze on the field' which is better than 'Graze of the cows field in'.
- Sentences that *correctly describes* the image content with higher *precision* have better *correctness* ratings. By precision, we focus on whether the sentence describes the 1) important *objects* 2) potential *actions* that are occurring and 3) the *location* of where the image is taken. E.g. '**Cows are grazing in the field**' is better than 'Cows are in the field' which is better than 'This is a field'.
- It is possible that a sentence is grammatically correct with high readability but does not relate to the image at all. E.g. 'Swans are flying above the pond' is perfectly readable but does not make sense in terms of correctness. Judge each scale independently.

Ratings are from a scale from **1** (lowest) to **5** (highest) which translates to the *level of English* of the writer whom you think produced this sentence :

Score	Readability	Correctness
1: Non speaker	Unreadable sentence with numerous grammar errors	Sentence content has no relevance to the image
2: Novice speaker	Barely readable sentence	Sentence content have only weak relevance to the image
3: Intermediate speaker	Reasonably readable sentence	Sentence content have some relevance to the image
4: Advance speaker	Readable sentence with few errors	Sentence content relates closely to image
5: Native speaker	Sentence with no grammatical errors	Sentence content relates perfectly to the image

Evaluation Result

Experiment	R_1 ,(length)	Relevance	Readability
Baseline 1, \mathcal{T}_{b1}^*	0.35,(8.2)	2.84 ± 1.40	3.64 ± 1.20
Baseline 2, \mathcal{T}_{b2}^*	0.39,(6.8)	2.14 ± 1.13	3.94 ± 0.91
HMM no corpus, \mathcal{T}_{eq}^*	0.42,(6.5)	2.44 ± 1.25	3.88 ± 1.18
Full HMM, \mathcal{T}^*	0.44,(6.9)	2.51 ± 1.30	4.10 ± 1.03
Human Annotation	0.68,(10.1)	4.91 ± 0.29	4.77 ± 0.42

Future Work

<ride>



<sit>



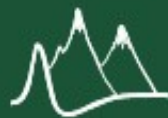
<fly>



Future Work

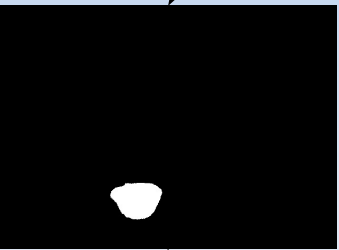
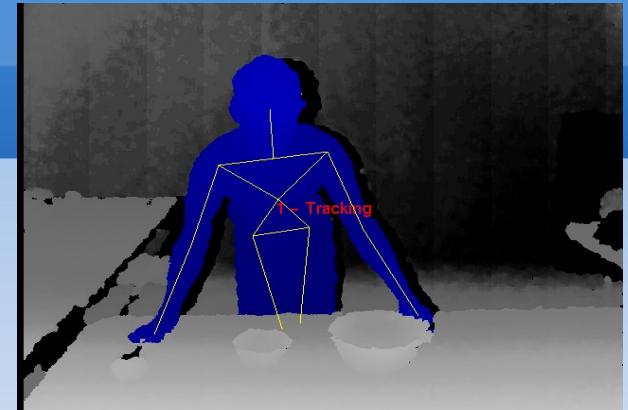
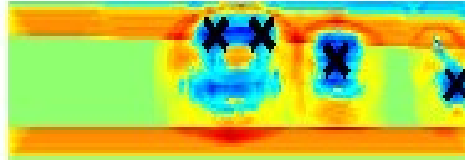
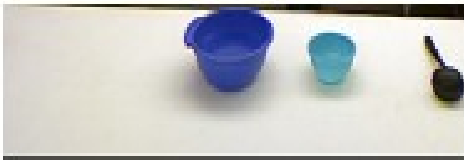


► Kinect

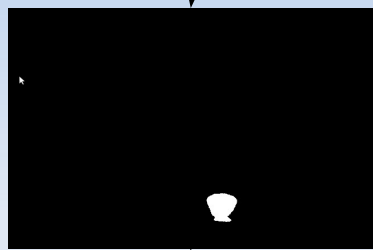


test image

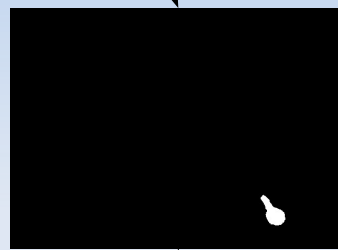
value map with extrema: white x is positive and black x is ne



Big Bowl



Small Bowl



Ladle

Pour

A person is using ladle to pour water into the bowl.

Thank You!



UMIACS

UNIVERSITY OF MARYLAND INSTITUTE FOR ADVANCED COMPUTER STUDIES