Segment-Phrase Table for Semantic Segmentation, Visual Entailment and Paraphrasing

Hamid Izadinia, Fereshteh Sadeghi,
Santosh K. Divvala, Hannaneh Hajishirzi,
Yejin Choi, Ali Farhadi

Presentated by Edward Banner

Outline

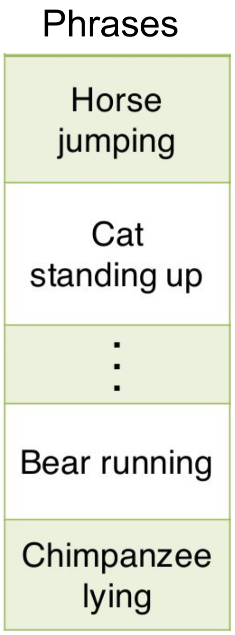
- What is a SPT?
- Motivation: What does a SPT enable us to do?
- How to build a SPT?
- How to make use of a SPT?
- Evaluation
- Discussion

What is a segment-phrase table?

One to many mapping from phrases
to segmentation models

What is a segment-phrase table?

One to many mapping from phrases to segmentation models



Phrases
Horse jumping
Cat standing up
⋮
Bear running
Chimpanzee lying

Image credit: Izadinia et al.

What is a segment-phrase table?

One to many mapping from phrases to segmentation models

Phrases	Segments
Horse jumping	
Cat standing up	
⋮	⋮
Bear running	
Chimpanzee lying	

Image credit: Izadinia et al.

Why build a segment-phrase table?

Many reasons!

Why build a segment-phrase table?

Entailment

If a horse is **grazing**, is it also **standing**?

Why build a segment-phrase table?

Entailment

If a horse is **grazing**, is it also **standing**?

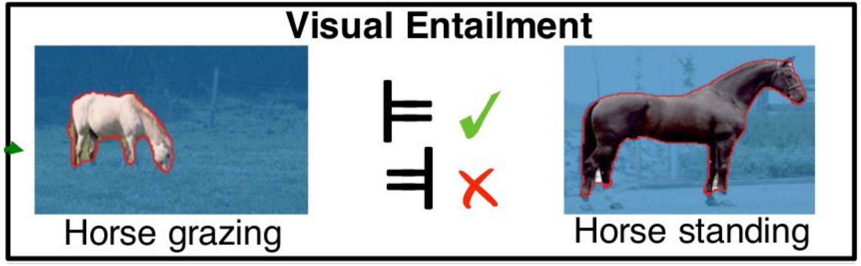


Image credit: Izadinia et al.

Why build a segment-phrase table?

Paraphrasing

Are “horse **jumping**” and “horse **leaping**” paraphrases of each other?

Why build a segment-phrase table?

Paraphrasing

Are “horse **jumping**” and “horse **leaping**” paraphrases of each other?

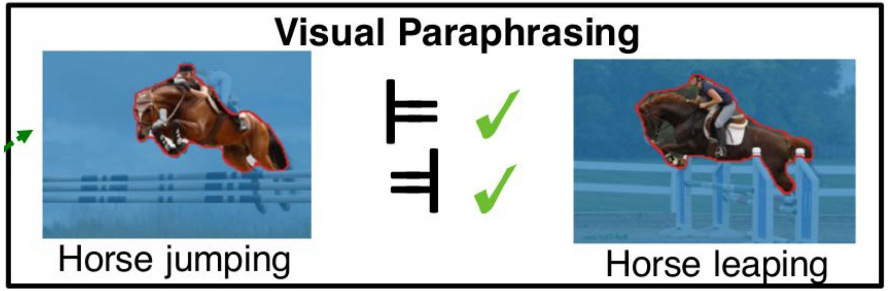


Image credit: Izadinia et al.

Why build a segment-phrase table?

Relative similarity

Is “**cat standing up**” closer to “**bear standing up**” or “**deer standing up**”?

Why build a segment-phrase table?

Relative similarity

Is “**cat standing up**” closer to “**bear standing up**” or “**deer standing up**”?

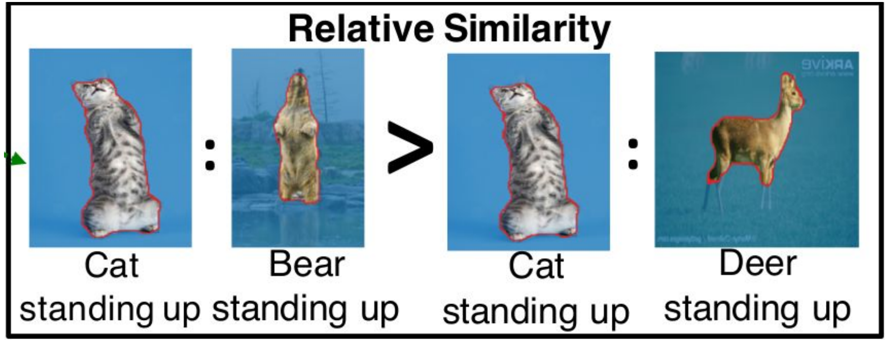


Image credit: Izadinia et al.

Why build a segment-phrase table?

Semantic segmentation

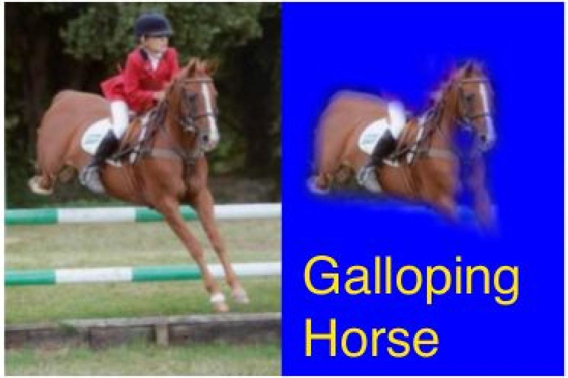
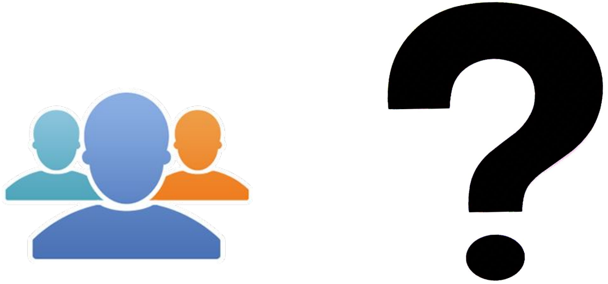


Image credit: Izadinia et al.

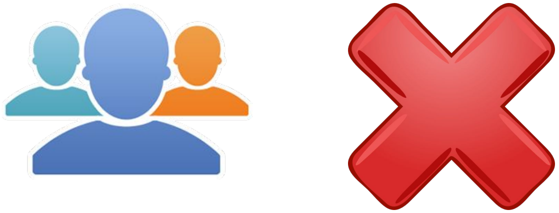
Considerations in building segment-phrase table

Human annotators?



Considerations in building segment-phrase table

Human annotators?



Too **expensive** to obtain human-labeled pixel labels

Opt instead for **weakly-supervised** approach instead

How do they build it?

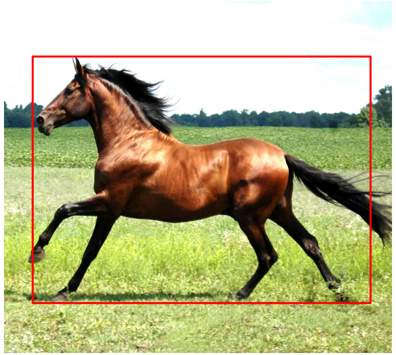
Three components:

1. Train a webly-supervised **detection model** *for each phrase*
2. Model each phrase as a **deformable parts model**
3. Learn **segmentation model** for each part

How do they build it?

1. Train a webly-supervised **detection model** *for each phrase*

e.g. **running horse**



How do they build it?

2. Model each phrase as a **deformable parts model**

Concerned about intra-class variation?

How do they build it?

2. Model each phrase as a **deformable parts model**

Concerned about intra-class variation?

horse

How do they build it?

2. Model each phrase as a **deformable parts model**

Concerned about intra-class variation?

horse

running horse

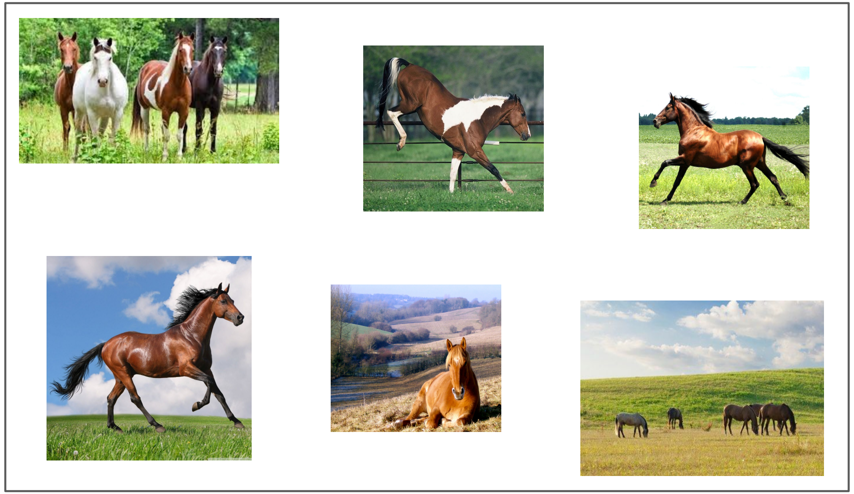
How do they build it?

2. Model each phrase as a **deformable parts model**

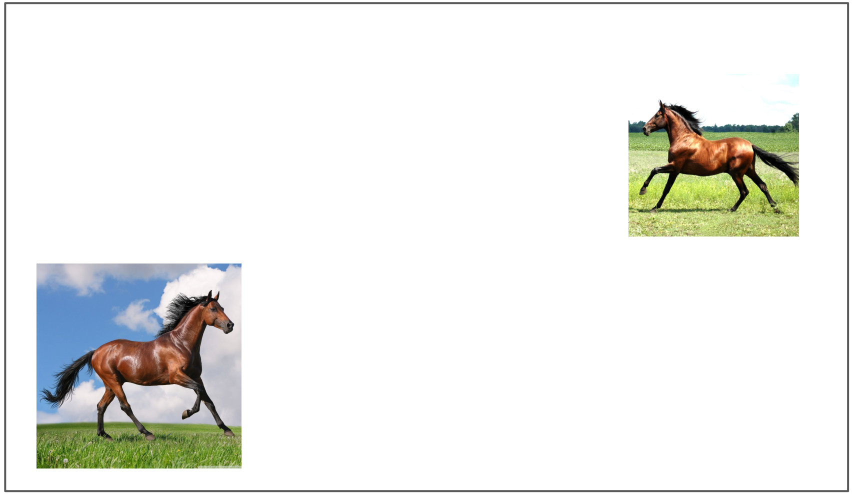
Concerned about intra-class variation?

Key insight: parts of phrases have low intra-class variation

horse



running horse



How do they build it?

$$E(\mathbf{x}) = \sum_{i \in V} u_i(x_i) + \sum_{(i,j) \in E} v_{ij}(x_i, x_j)$$

3. Learn **segmentation model** $\theta_c^{fg}, \theta_c^{bg}$ for each part

Model superpixels with GMM and solve with EM and Graphcut

Rough initialization with Grabcut and HOG root filter

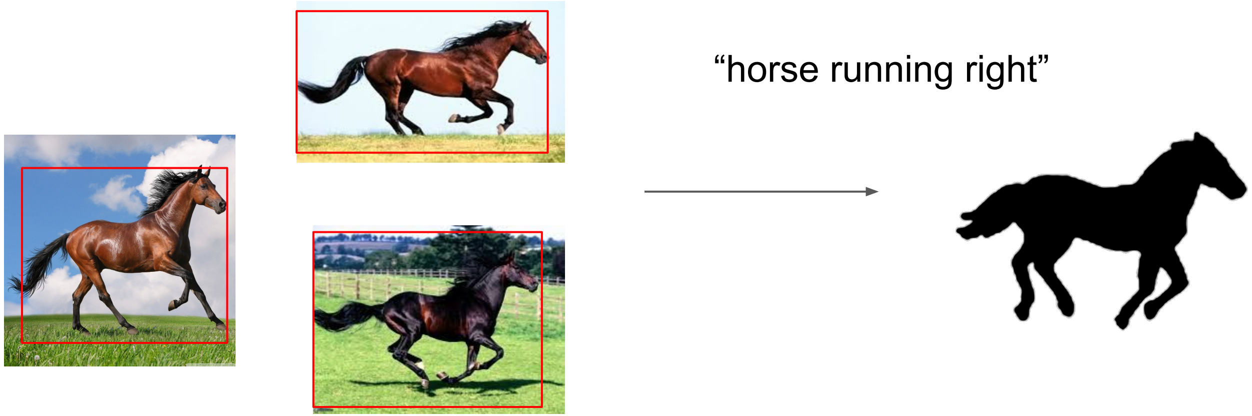
$$E(\mathbf{x}) = \sum_{i \in V} u_i(x_i) + \sum_{(i,j) \in E} v_{ij}(x_i, x_j)$$

How do they build it?

3. Learn **segmentation model** $\theta_c^{fg}, \theta_c^{bg}$ for each part

Model superpixels with GMM and solve with EM and Graphcut

Rough initialization with Grabcut and HOG root filter



“horse running right”

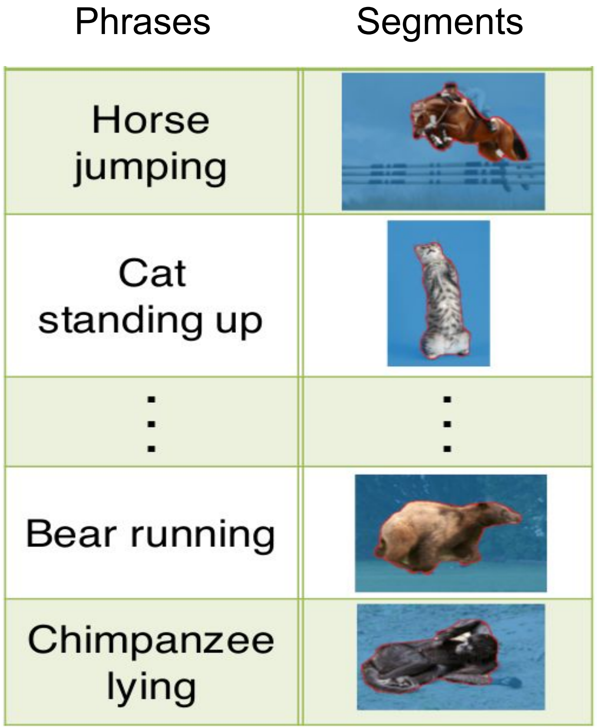
Segment-phrase table built

Results:

For each phrase, we have learned:

- Bounding box detector
- Segmentation model for each part

What can we do now?



Phrases	Segments
Horse jumping	
Cat standing up	
⋮	⋮
Bear running	
Chimpanzee lying	

Image credit: Izadinia et al.

Semantic segmentation

Example: “horse”

Test image

Image credit: Izadinia et al.

Semantic segmentation

Example: “horse”

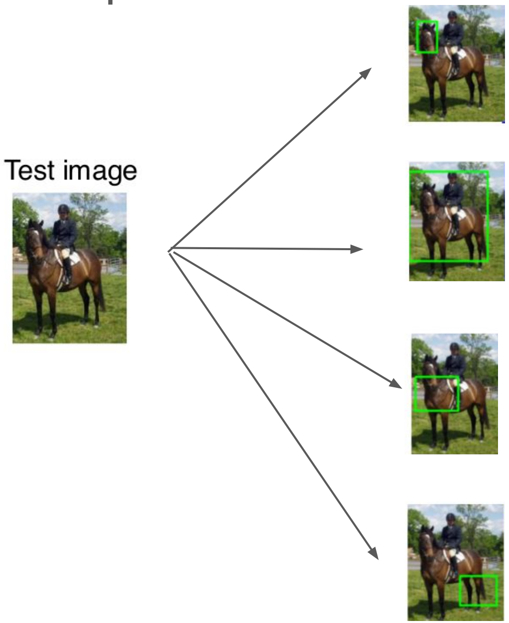


Image credit: Izadinia et al.

Semantic segmentation

Example: “horse”

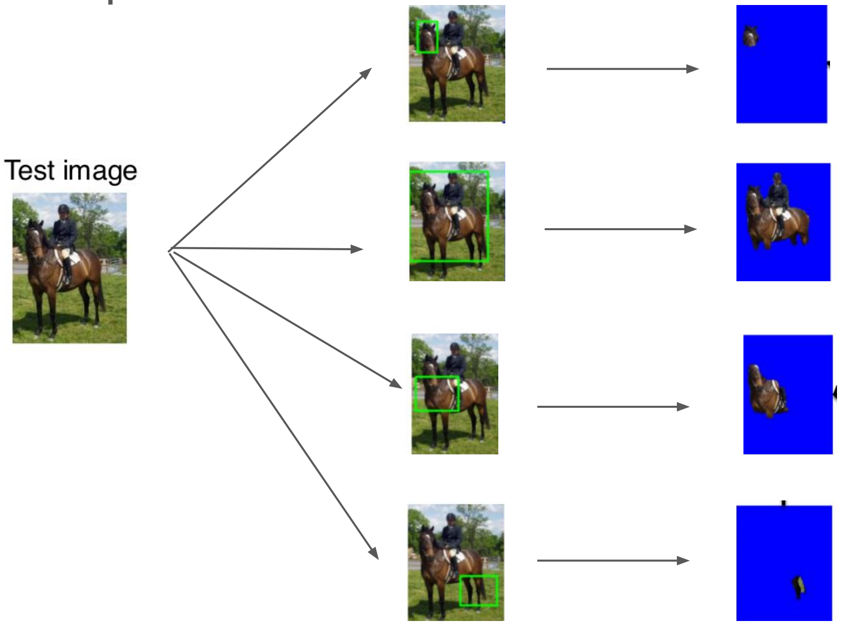


Image credit: Izadinia et al.

Semantic segmentation

Example: “horse”

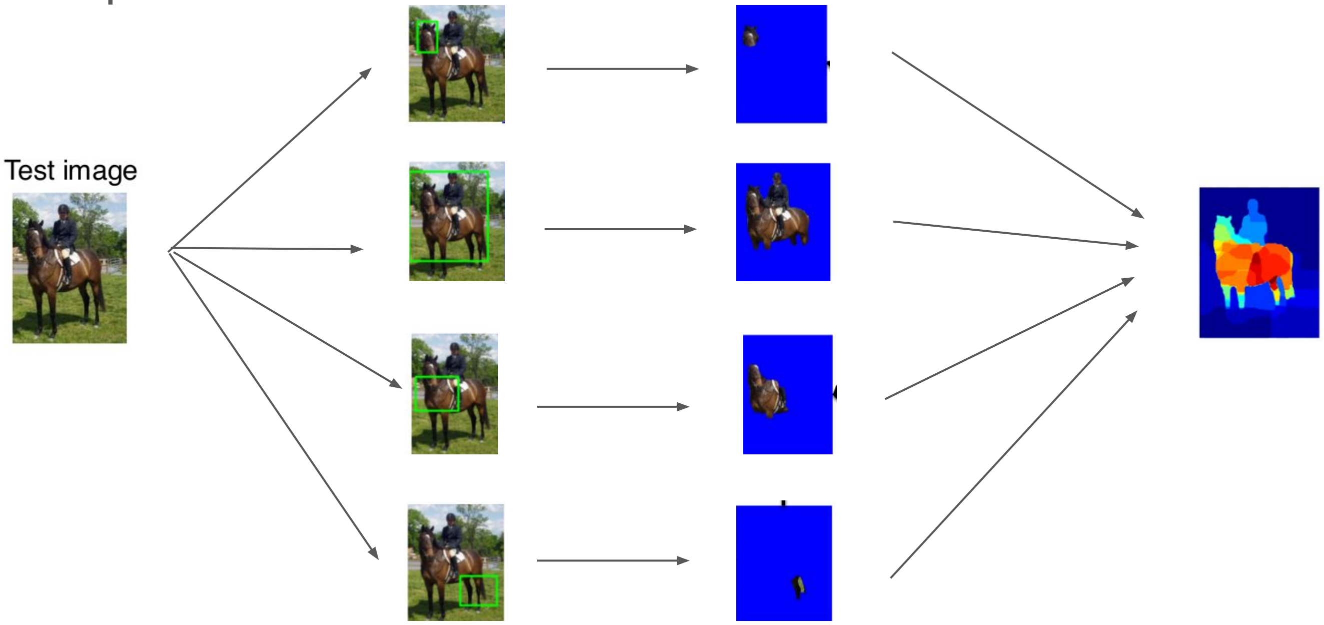


Image credit: Izadinia et al.

Semantic segmentation

Example: “horse”

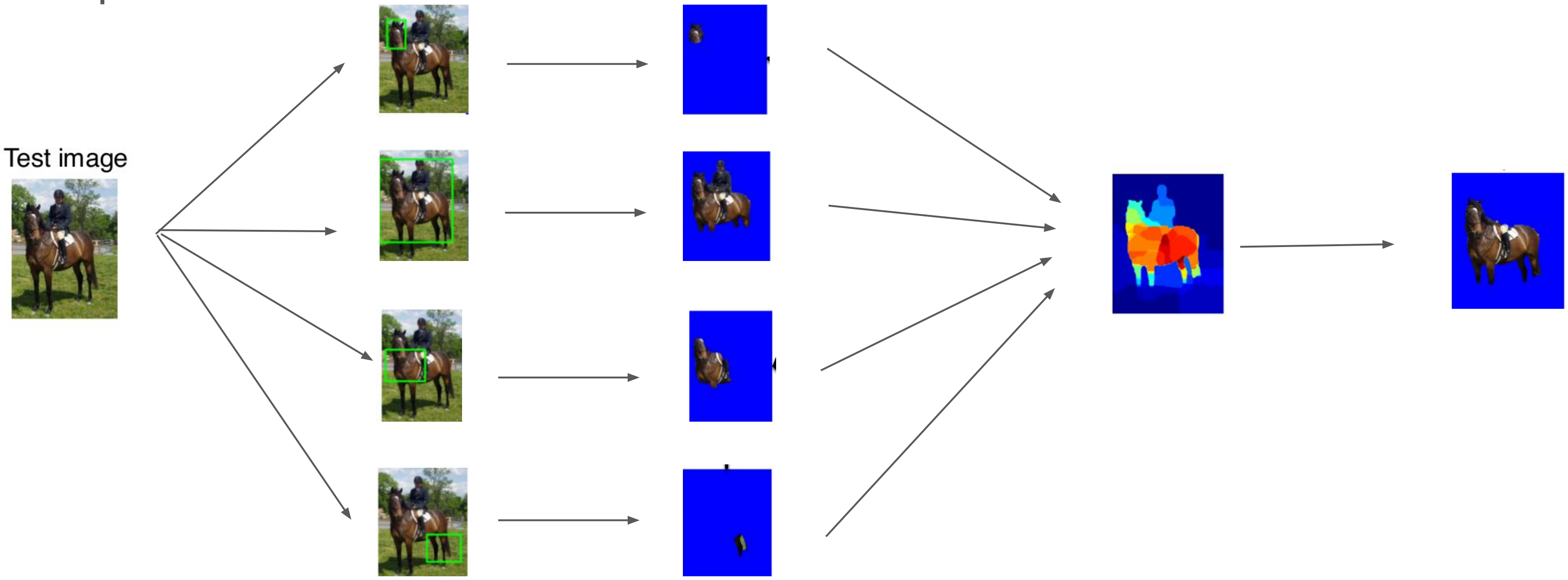


Image credit: Izadinia et al.

Semantic segmentation *using linguistic constraints*

Example: “horse”

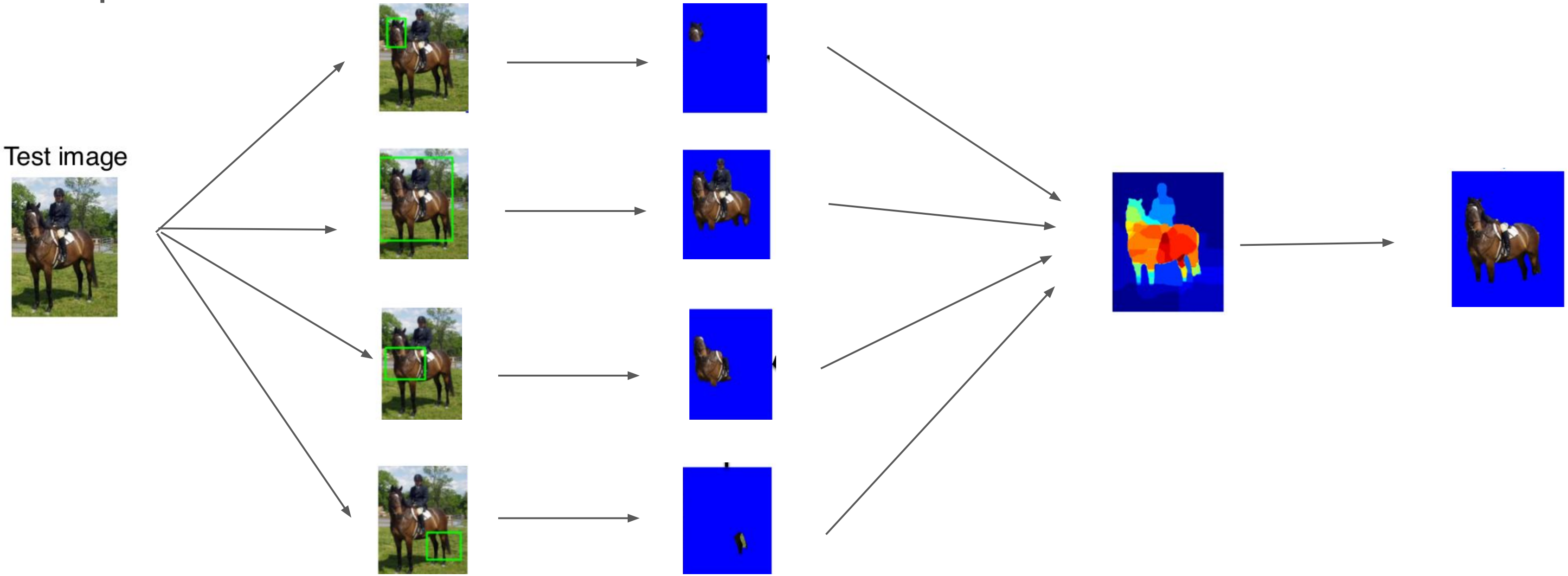


Image credit: Izadinia et al.

Semantic segmentation *using linguistic constraints*

Example: “horse”

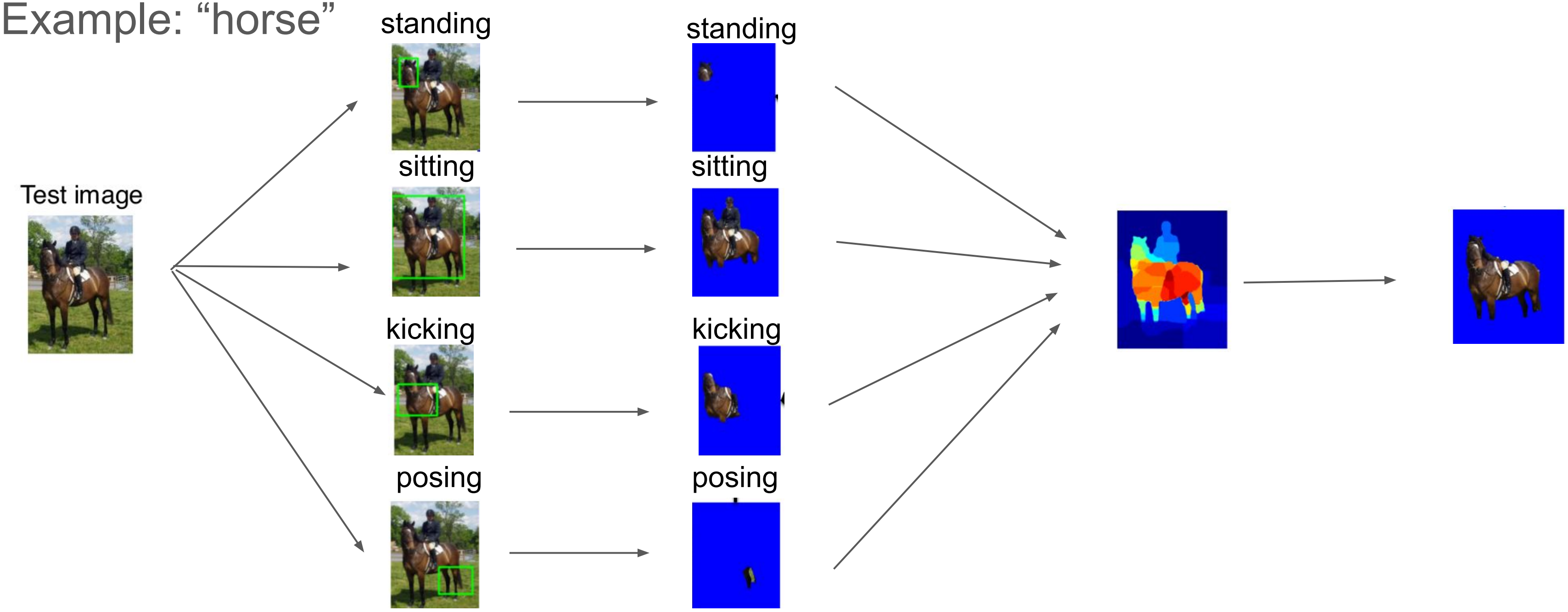


Image credit: Izadinia et al.

Semantic segmentation *using linguistic constraints*

Example: “horse”

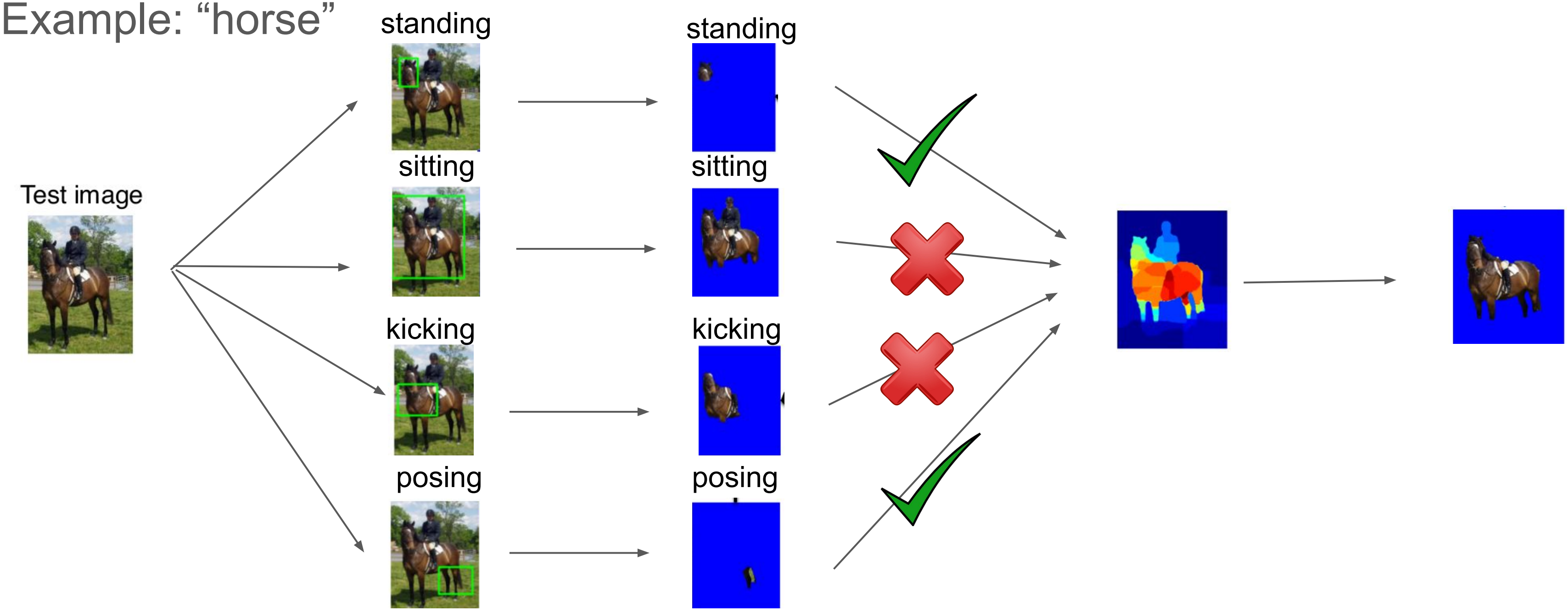


Image credit: Izadinia et al.

Entailment

$$\text{entail}(X \vDash Y) := Sim_{\overrightarrow{R2I}}(X, Y) - Sim_{\overrightarrow{R2I}}(Y, X),$$

Does phrase X entail phrase Y?

Intuition: All segments for which phrase X
is a valid description, then phrase Y is also a valid description

Entailment

$$\text{entail}(X \vDash Y) := Sim_{\overrightarrow{R2I}}(X,Y) - Sim_{\overrightarrow{R2I}}(Y,X),$$

Does phrase X entail phrase Y?

Intuition: All segments for which phrase X is a valid description, then phrase Y is also a valid description

horse grazing

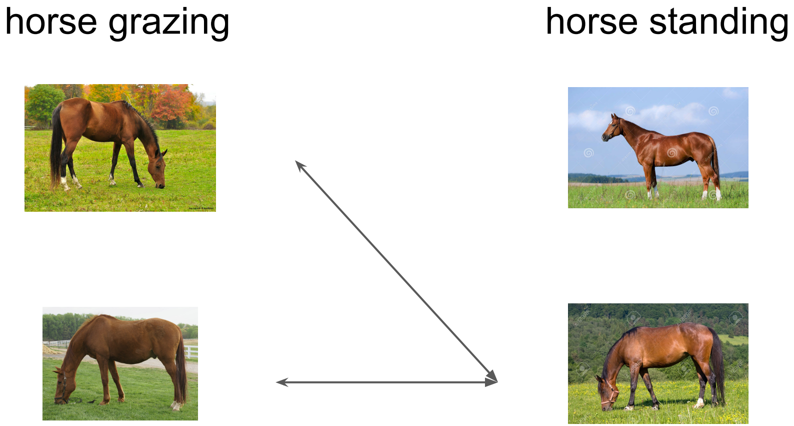
horse standing

Entailment

$$\mathrm{entail}(X \models Y) := Sim_{\overrightarrow{R2I}}(X, Y) - Sim_{\overrightarrow{R2I}}(Y, X),$$

Does phrase X entail phrase Y?

Intuition: All segments for which phrase X is a valid description, then phrase Y is also a valid description


horse grazing

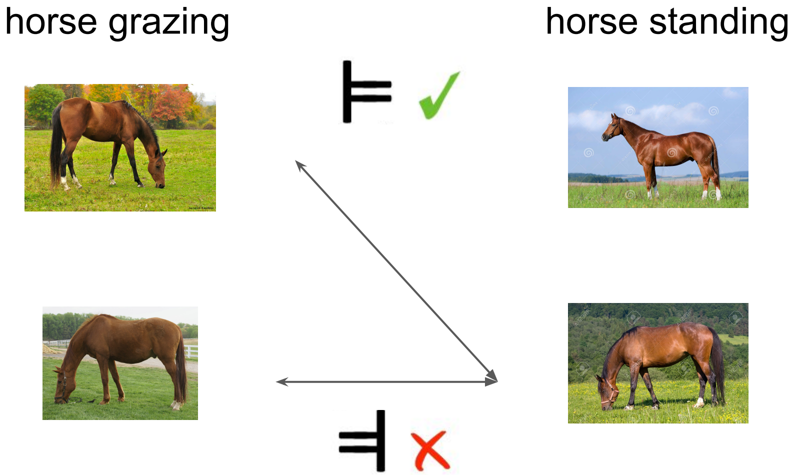
horse standing

Entailment

$$\text{entail}(X \vDash Y) := Sim_{\overrightarrow{R2I}}(X,Y) - Sim_{\overrightarrow{R2I}}(Y,X),$$

Does phrase X entail phrase Y?

Intuition: All segments for which phrase X is a valid description, then phrase Y is also a valid description


horse grazing

horse standing

⊨ ✓

⫤ ✗

Paraphrasing

Are phrase X and phrase Y paraphrases of each other?

Strategy: compute X ⊨ Y and Y ⊨ X and say they're paraphrases if they're close

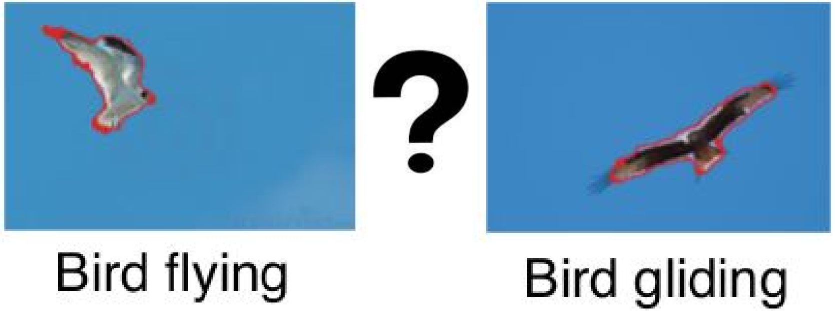


Image credit: Izadinia et al.

Paraphrasing

Are phrase X and phrase Y paraphrases of each other?

Strategy: compute X ⊨ Y and Y ⊨ X and say they’re paraphrases if they’re close

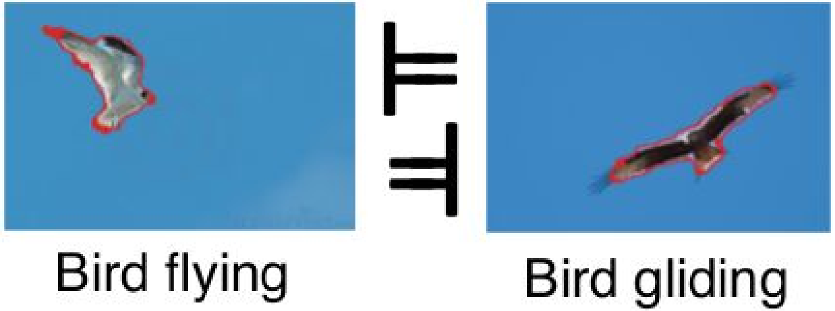


Image credit: Izadinia et al.

Relative Semantic Similarity

Is phrase X closer to phrase Y or phrase Z?

Strategy: compute $X \models Y$ and $X \models Z$ and pick highest number of the two

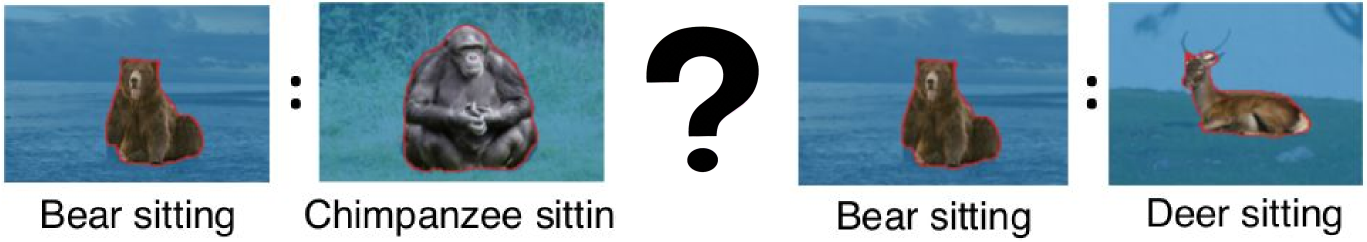


Image credit: Izadinia et al.

Relative Semantic Similarity

Is phrase X closer to phrase Y or phrase Z?

Strategy: compute X ⊨ Y and X ⊨ Z and pick highest number of the two

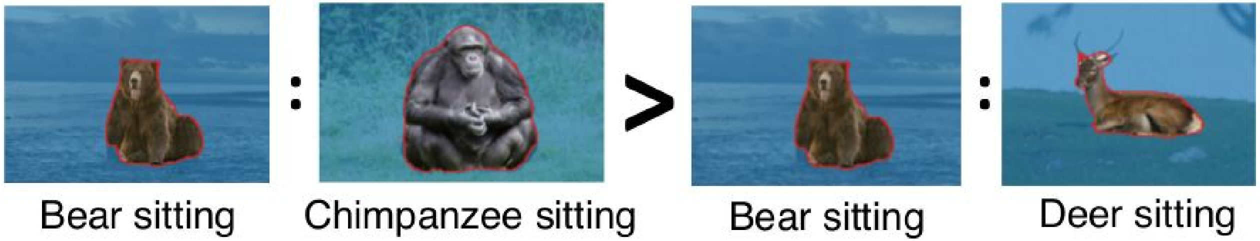


Image credit: Izadinia et al.

Evaluation - Takeaways

Semantic segmentation state of the art or near it

Highlights tradeoffs between unsupervised approach on large data and supervised approaches on small dataset

Linguistic constraints help semantic segmentation

SPT approach beats language-only and vision-only baselines on entailment, paraphrasing, and relative similarity

Discussion

Discussion

Leverage supervision

Variable number of part models per phrase

Larger evaluation dataset

Comparison against state-of-the-art entailment and paraphrase systems