



Learning Models for Object Recognition from Natural Language Descriptions

Josiah Wang
Katja Markert
Mark Everingham

School of Computing
University of Leeds

Main Idea

- Learn models using **only** textual descriptions
 - **No** training images used

Monarch *Danaus plexippus*

+ SAVE TO LIST

1 SEND ECARD



Family: Nymphalidae, Brush-footed Butterflies [view all from this family](#)

Description 3 1/2-4" (89-102 mm). Very large, with FW long and drawn out. Above, bright, burnt-orange with black veins and black margins sprinkled with white dots; FW tip broadly black interrupted by larger white and orange spots. Below, paler, duskier orange. 1 black spot appears between HW cell and margin on male above and below. Female darker with black veins smudged.

(Source: eNature.com)

Main Idea

- Conventional approaches require many training images
 - Difficult to scale to large number of categories
- Related work in CVPR 2009
 - Farhadi *et al.* (2009) & Lampert *et al.* (2009)
 - Describe object categories using named attributes
 - Attributes are defined by hand for categories
 - We learn these from textual descriptions

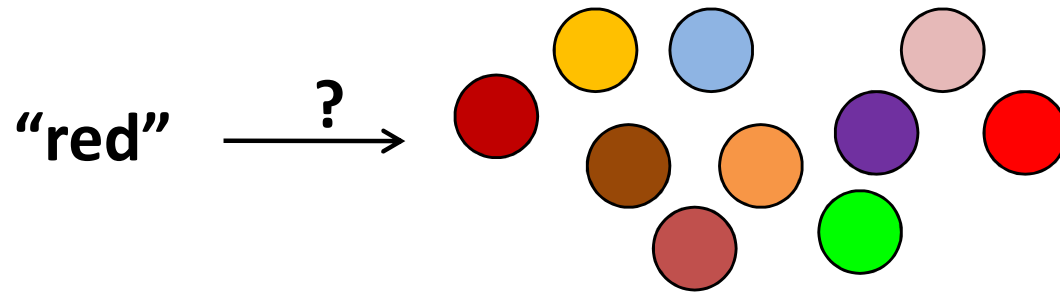
Textual Descriptions

- Define **appearance** properties of an object category
- Readily available for certain object categories (butterflies, flowers, sign language, judo moves, *etc.*)

Very large, with forewing long and drawn out. Above, **bright, burnt-orange** with **black veins** and **black margins** sprinkled with **white dots**; forewing **tip broadly black** interrupted by larger **white and orange spots**. Below, paler, duskier **orange**.

Challenges

- Mapping between text and images



- Extracting information from textual descriptions
 - Parsing
- Short descriptions
- Some described properties are not visible in images

Dataset

- Ten butterfly categories
- Training set: Textual descriptions **only** (from eNature)
- Test set: Butterfly images (from Google Images)



*Danaus
plexippus*



*Heliconius
charitonius*



*Heliconius
erato*



*Junonia
coenia*



*Lycaena
phlaeas*



*Nymphalis
antiopa*



*Papilio
cresphontes*



*Pieris
rapae*



*Vanessa
atalanta*



*Vanessa
cardui*

Method Outline

training descriptions

Above, bright, burnt-orange with black veins and black margins sprinkled with white dots; FW tip broadly black interrupted by larger white and orange spots.

Natural Language Processing (NLP)

Representation from text

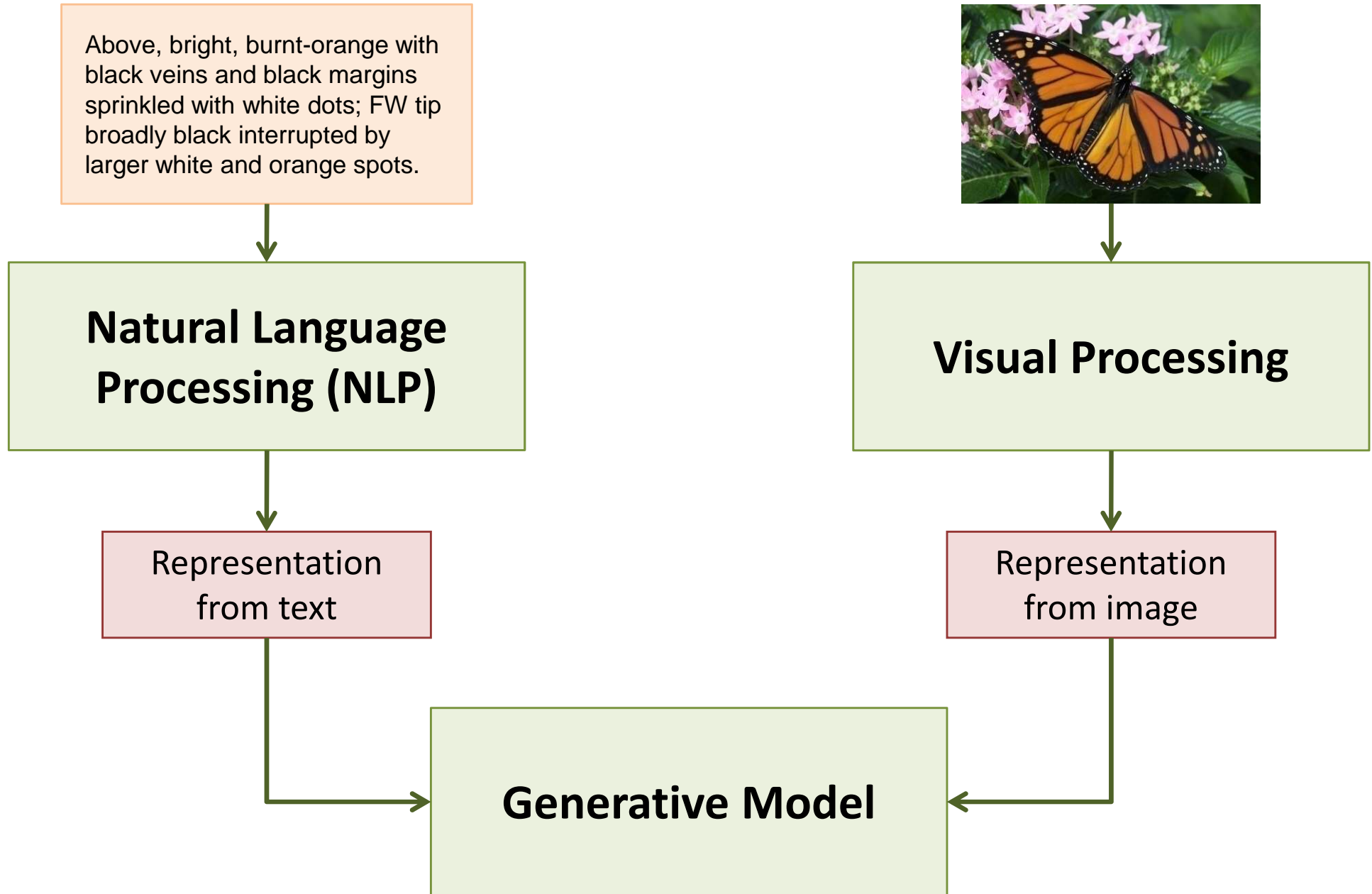
Generative Model

test image

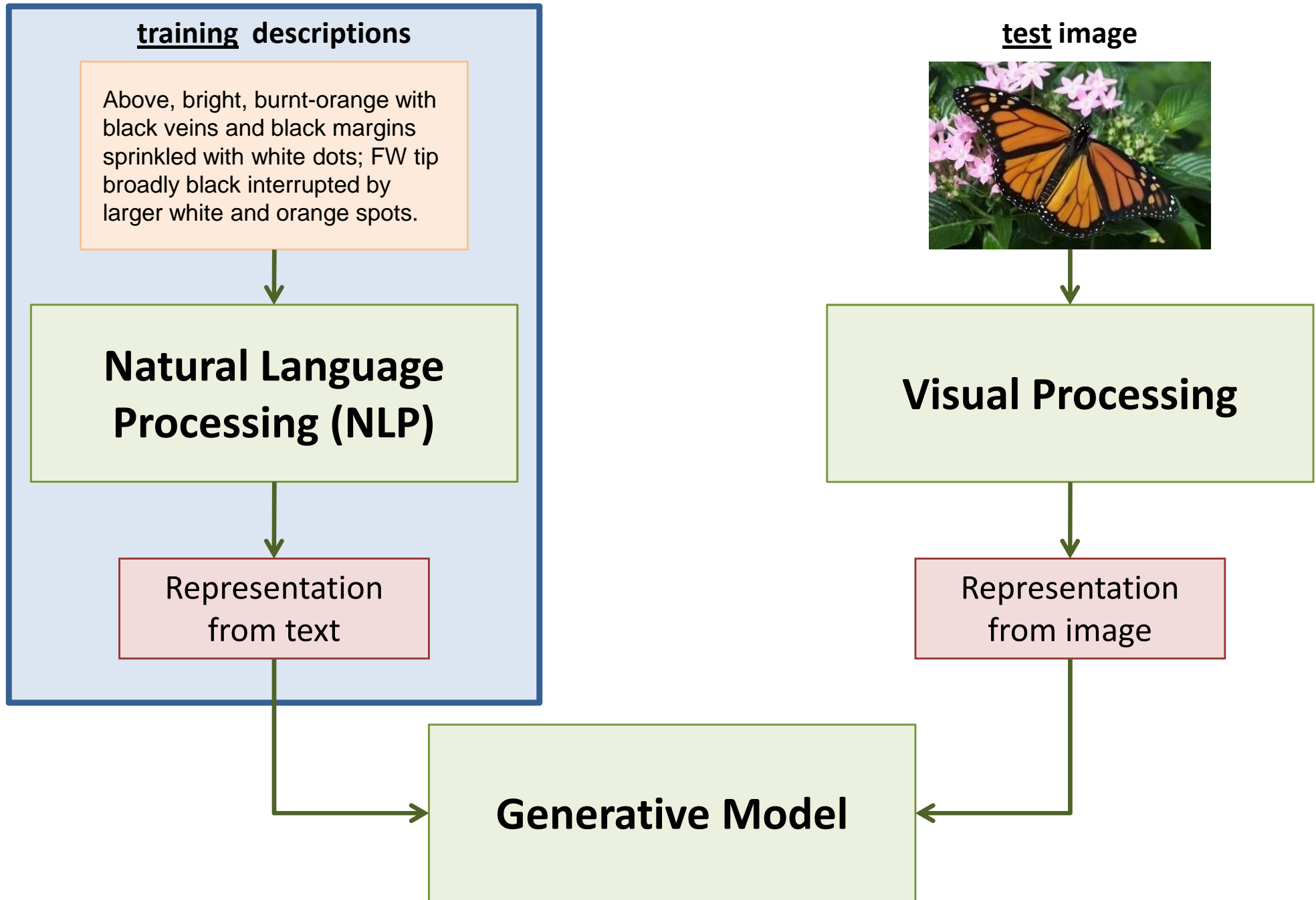


Visual Processing

Representation from image



Natural Language Processing (NLP)



Natural Language Processing (NLP)

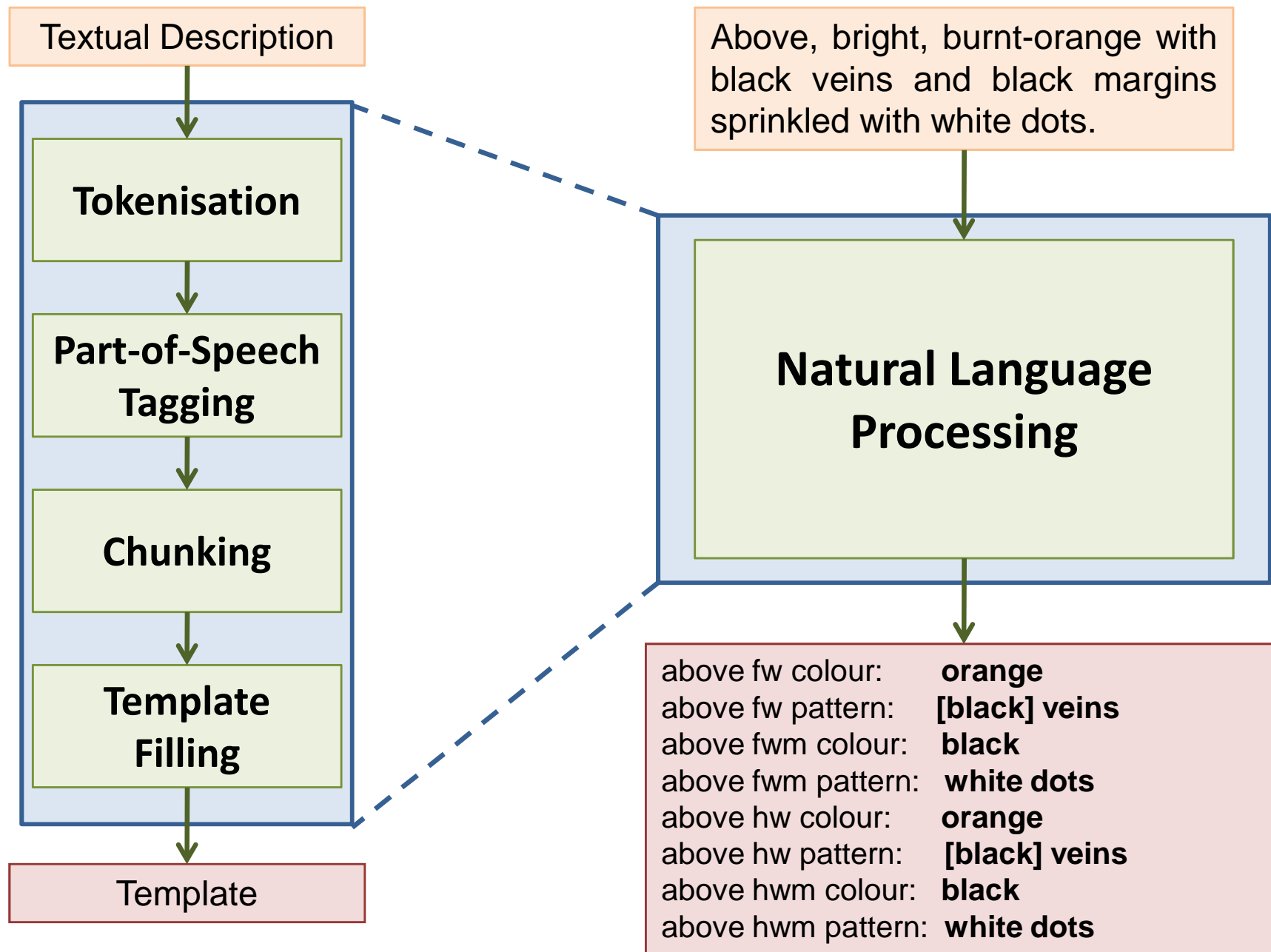
- Convert descriptions into **templates**

Above, bright, burnt-**orange** with **black veins** and **black margins** sprinkled with **white dots**; forewing tip broadly **black** interrupted by larger **white and orange spots**. Below, paler, duskier **orange**.

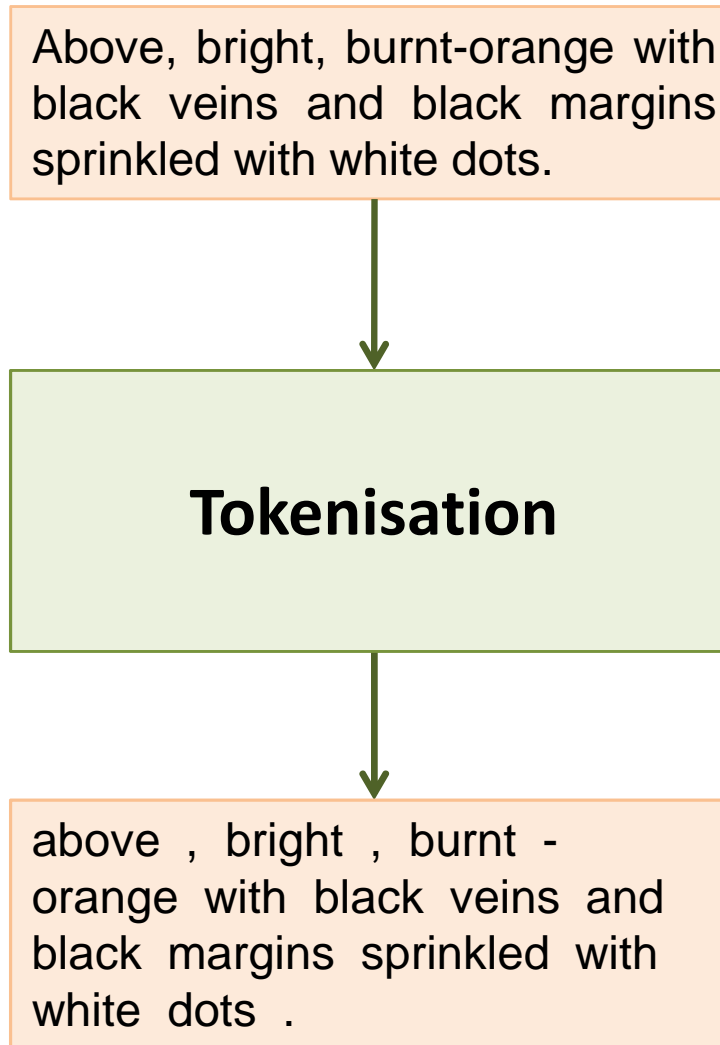
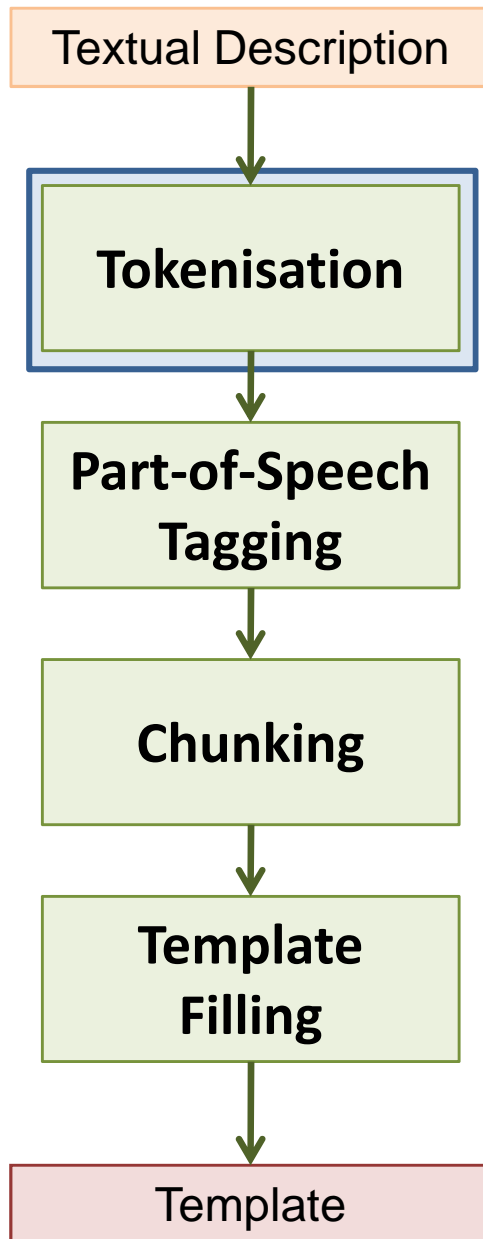
**Natural Language
Processing**

above fw colour	: orange
above fw pattern	: [black] veins
above fwm colour	: black
above fwm pattern	: [white] dots
	: [white and orange] spots
above hw colour	: orange
above hw pattern	: [black] veins
above hwm colour	: black
above hwm pattern	: [white] dots
below fw colour	: orange
below fw pattern	:
below fwm colour	:
below fwm pattern	:
below hw colour	: orange
below hw pattern	:
below hwm colour	:
below hwm pattern	:

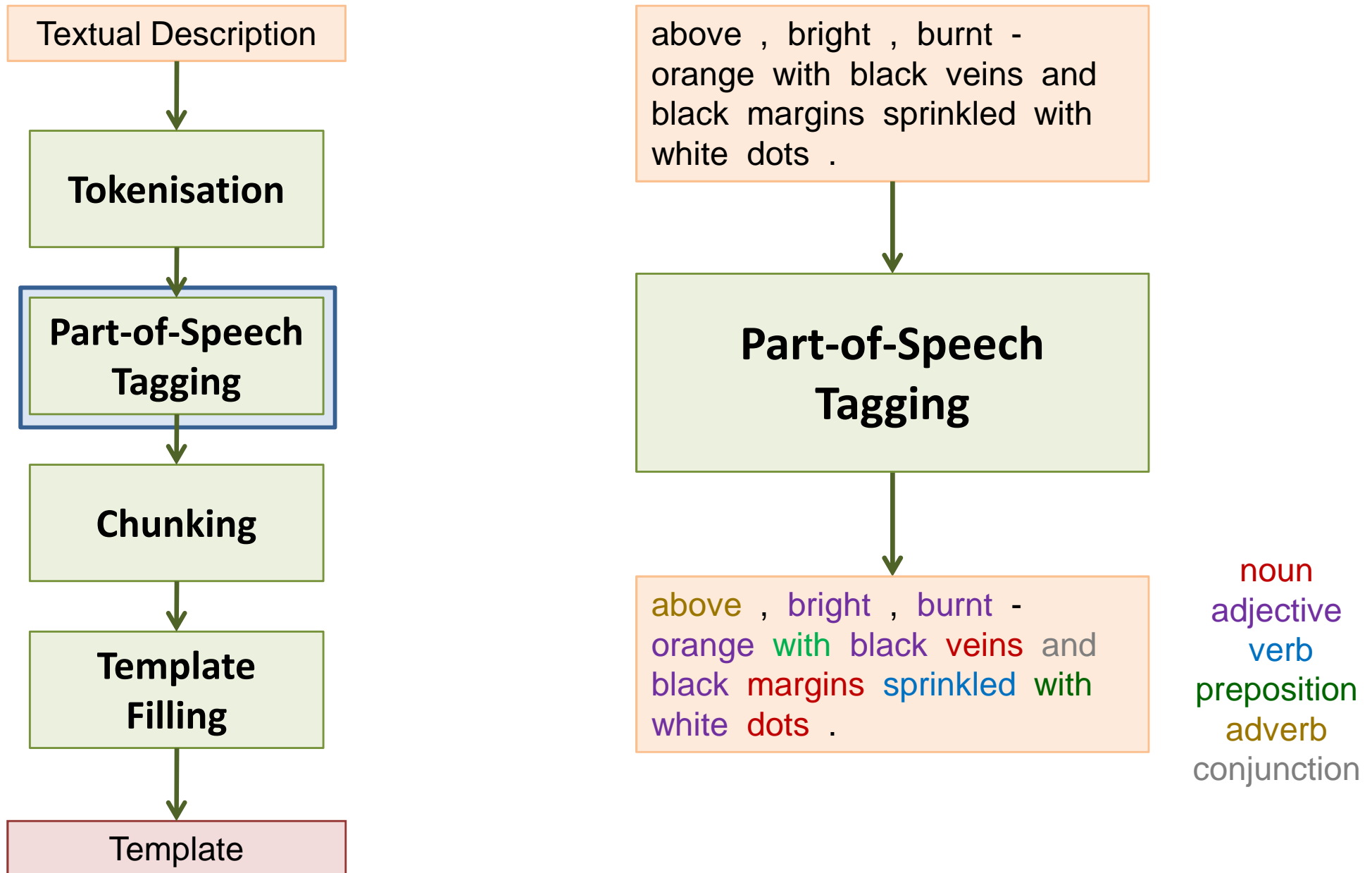
Natural Language Processing (NLP)



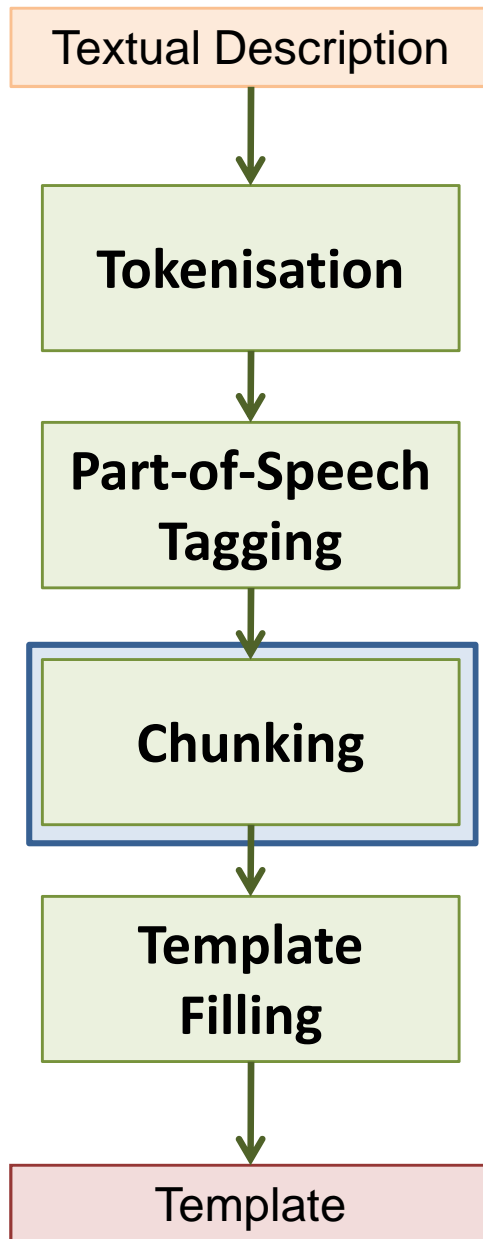
Natural Language Processing (NLP)



Natural Language Processing (NLP)



Natural Language Processing (NLP)



above , bright , burnt -
orange with black veins and
black margins sprinkled with
white dots .

noun
adjective
verb
preposition
adverb
conjunction

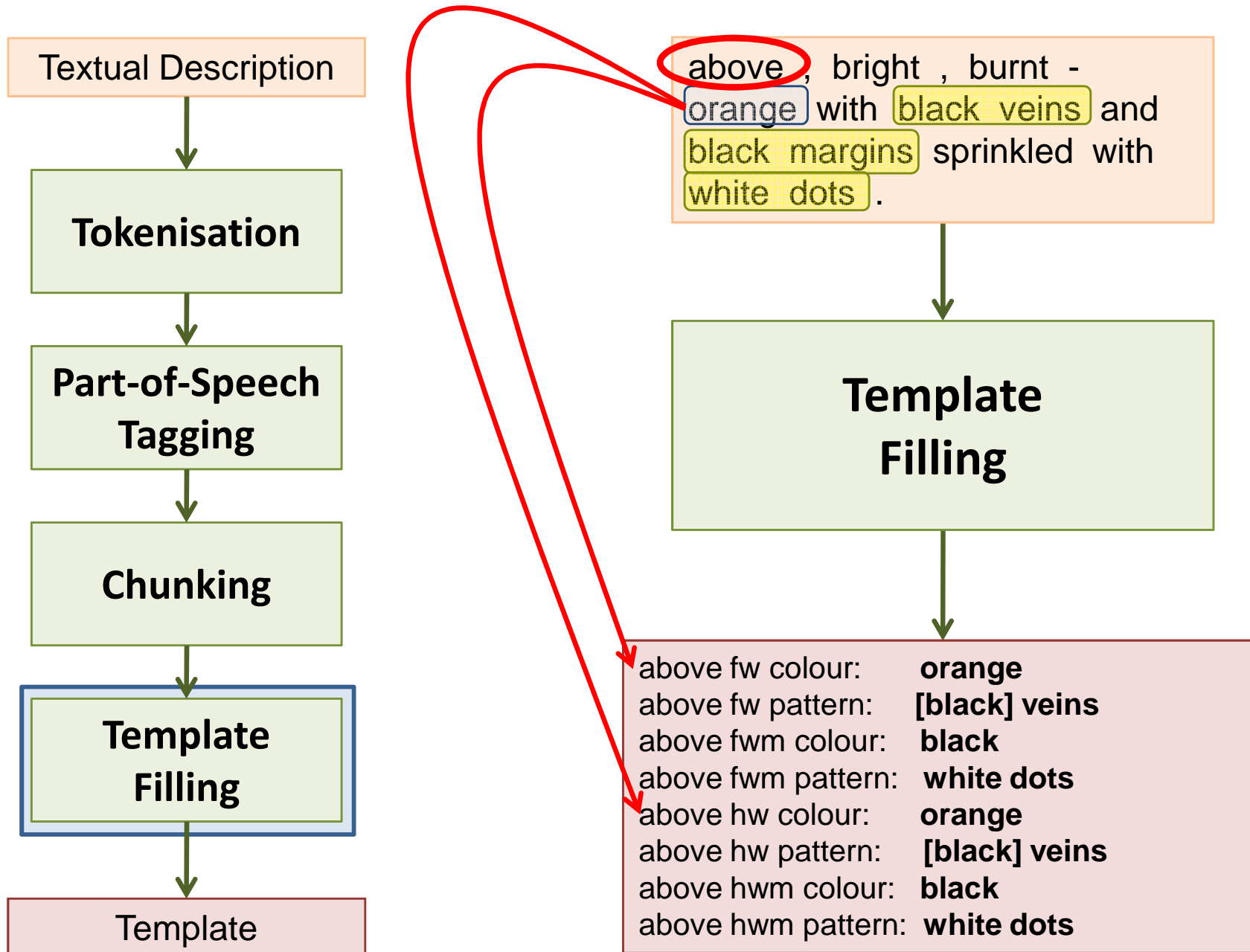


above , bright , burnt -
orange with black veins and
black margins sprinkled with
white dots .

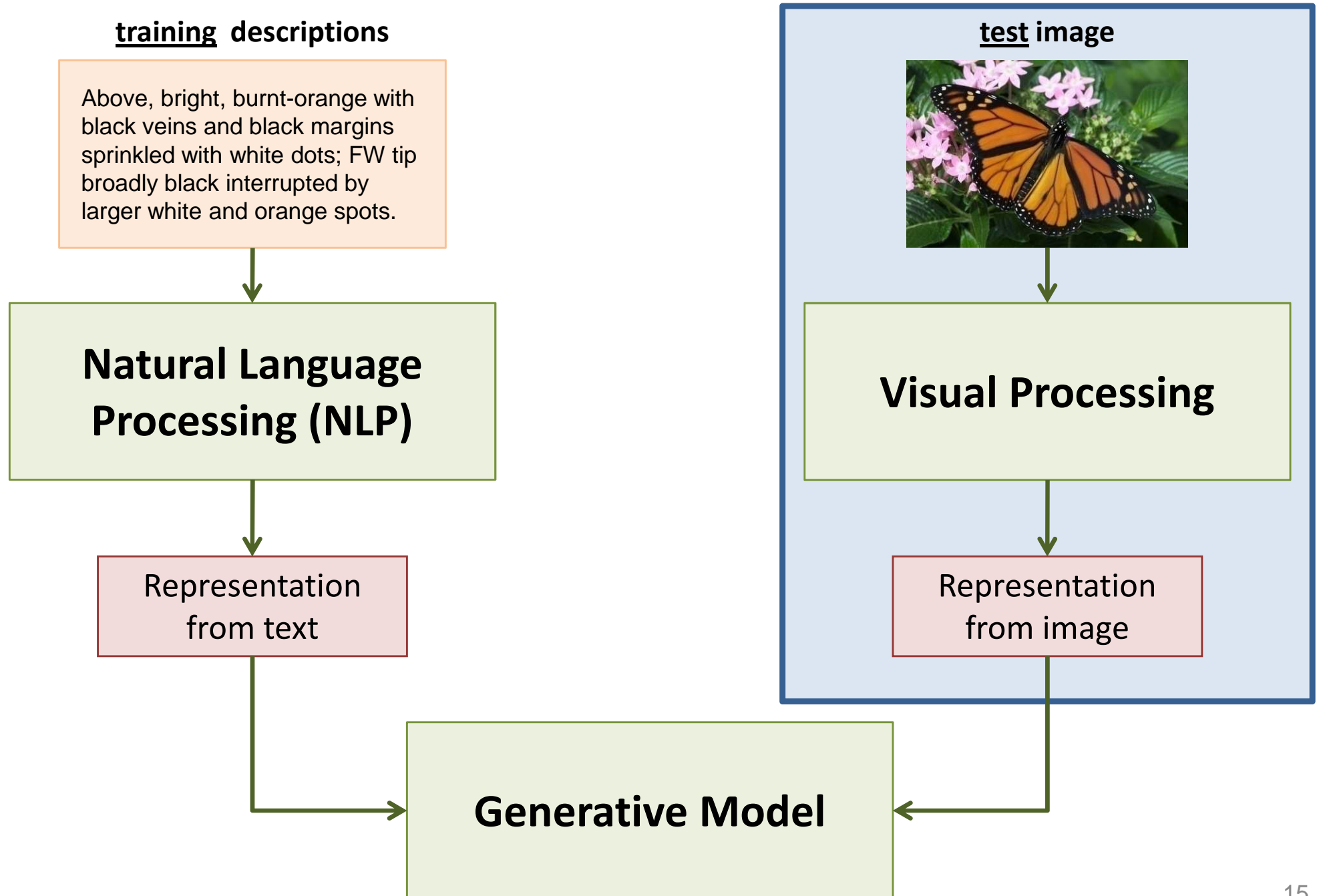
adjective phrase

noun phrase

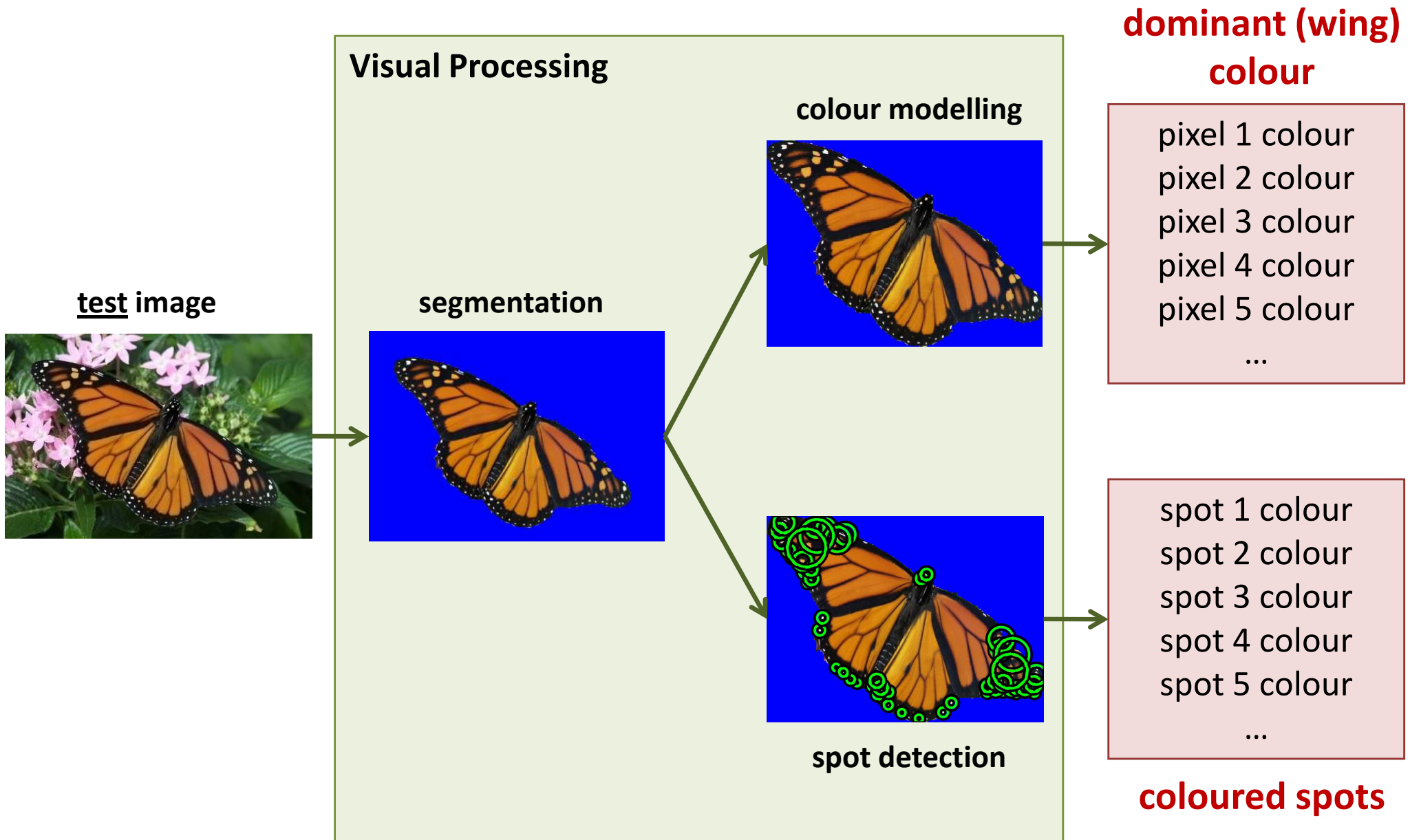
Natural Language Processing (NLP)



Visual Processing

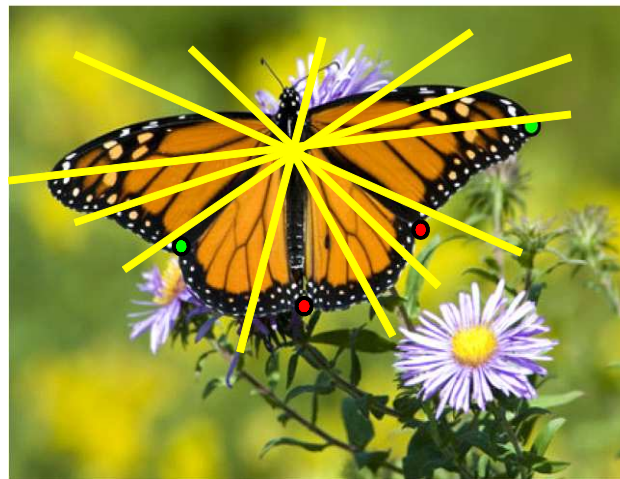
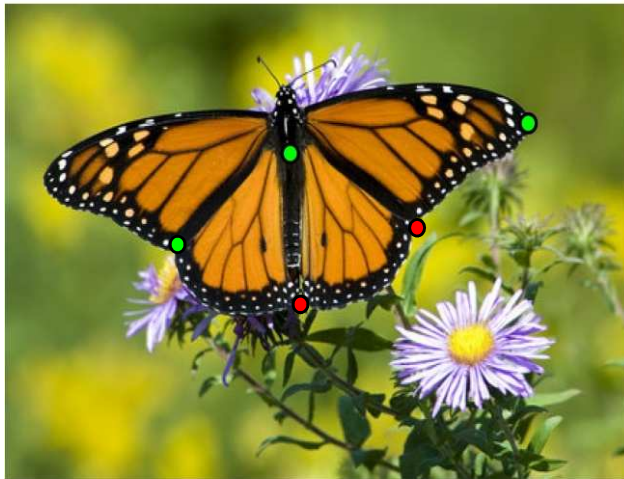


Visual Processing



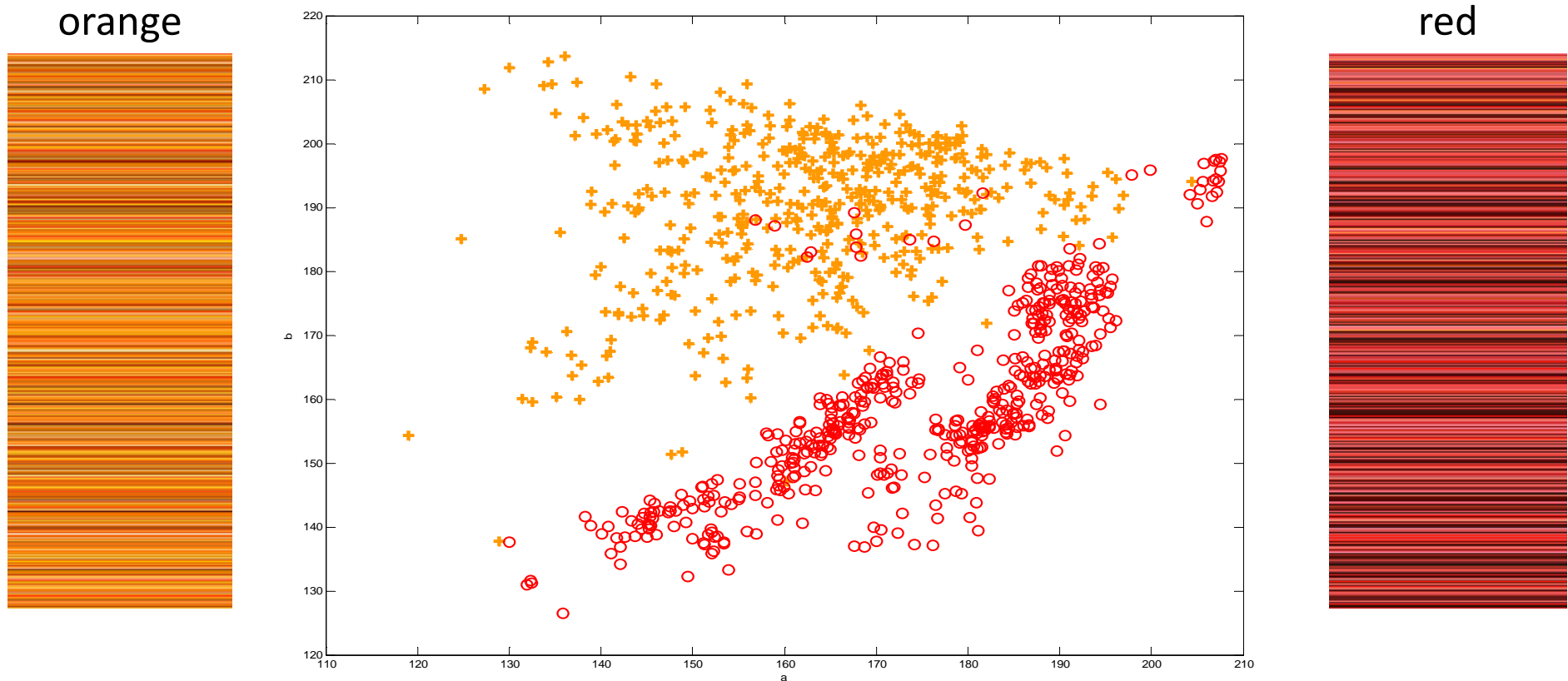
Visual Processing

- Segmentation
 - Interactive ‘star shape’ graph-cut (Veksler 2008)
 - User selects a centre for the butterfly
 - User may specify more foreground/background points



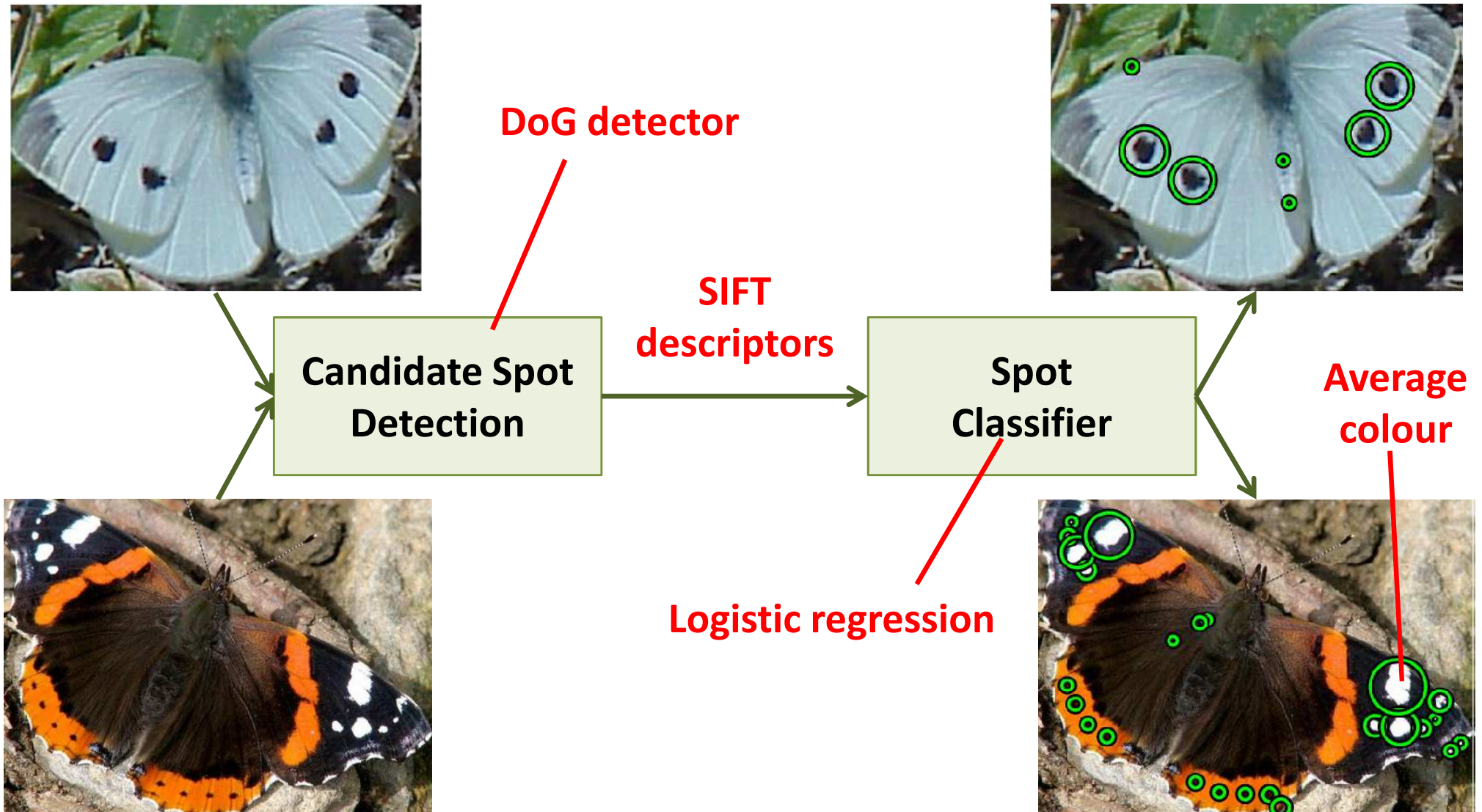
Visual Processing

- Colour modelling
 - Relate a colour name *e.g.* “orange” to $L^*a^*b^*$ values
 - Learn Parzen density model from selected pixels in butterfly images (category labels are **not** used)



Visual Processing

- Spot detection



Generative Model

training descriptions

Above, bright, burnt-orange with black veins and black margins sprinkled with white dots; FW tip broadly black interrupted by larger white and orange spots.

test image



Natural Language Processing (NLP)

Visual Processing

Representation from text

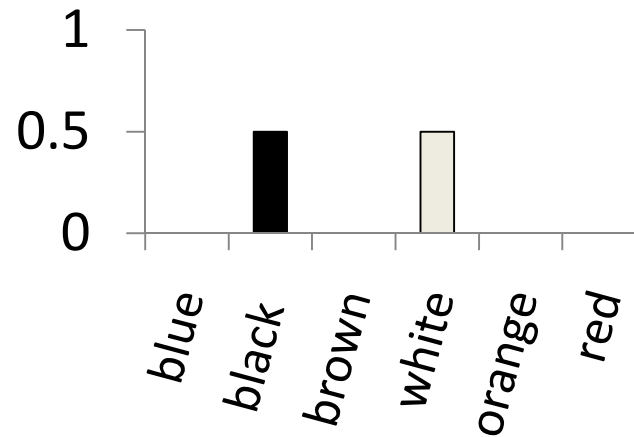
Representation from image

Generative Model

Generative Model

- Template → Spot Colour Name Prior

above fw colour : black
above fw pattern : [orange] bars
above fwm pattern : [white] spots
above hw colour : black
above hwm pattern : [blue] patch
above hwm pattern : [black] spots

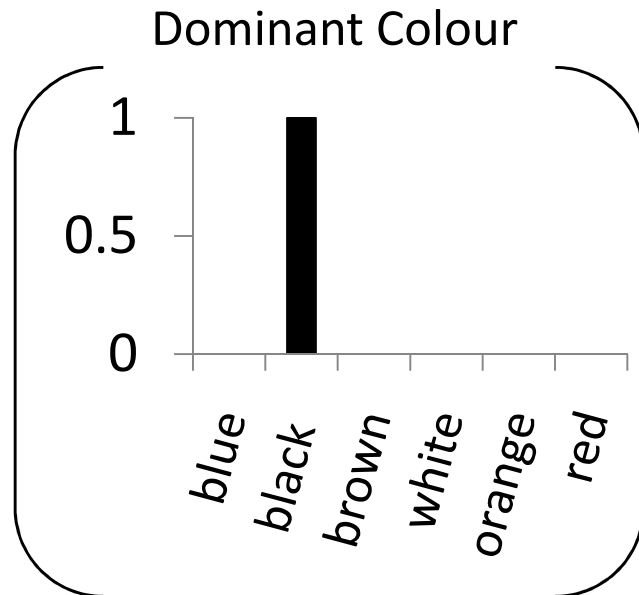


Spot Colour Name Prior

Generative Model

- Template → Wing Colour Name Prior

above fw colour	· black
above fw pattern	: [orange] bars
above fwm pattern	: [white] spots
above hw colour	· black
above hwm pattern	: [blue] patch
above hwm pattern	: [black] spots

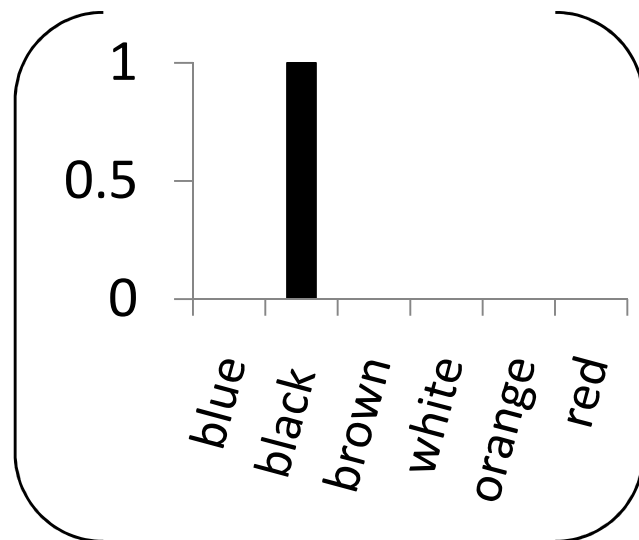


Generative Model

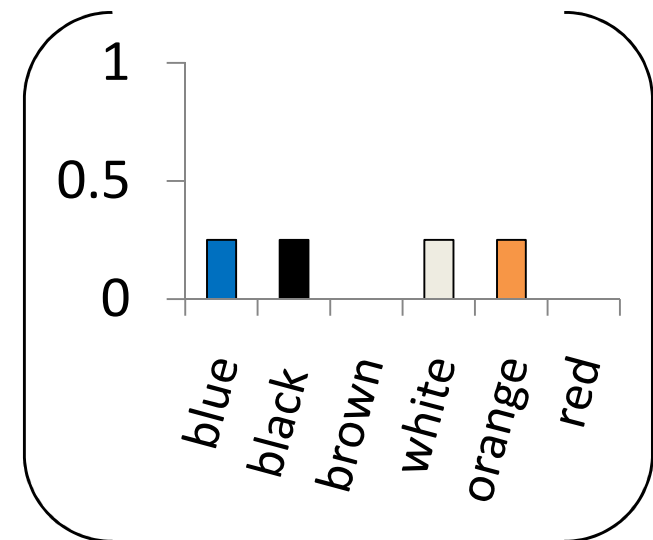
- Template → Wing Colour Name Prior

above fw colour : black
above fw pattern : [orange] bars
above fwm pattern : [white] spots
above hw colour : black
above hwm pattern : [blue] patch
above hwm pattern : [black] spots

Dominant Colour

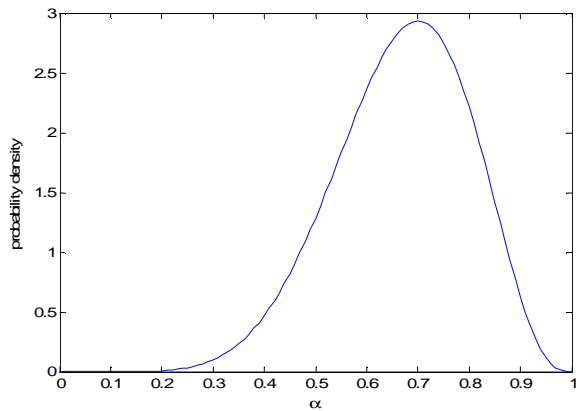


'Other' Colour



Generative Model

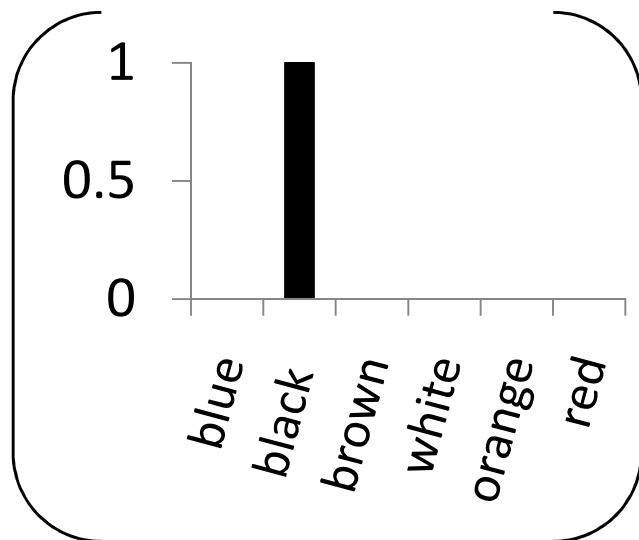
- Template \rightarrow Wing Colour Name Prior



above fw colour : black
above fw pattern : [orange] bars
above fwm pattern : [white] spots
above hw colour : black
above hwm pattern : [blue] patch
above hwm pattern : [black] spots

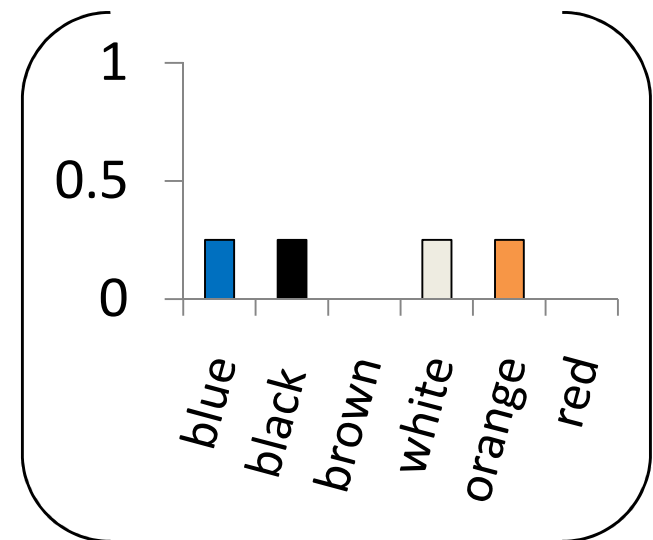
α

Dominant Colour



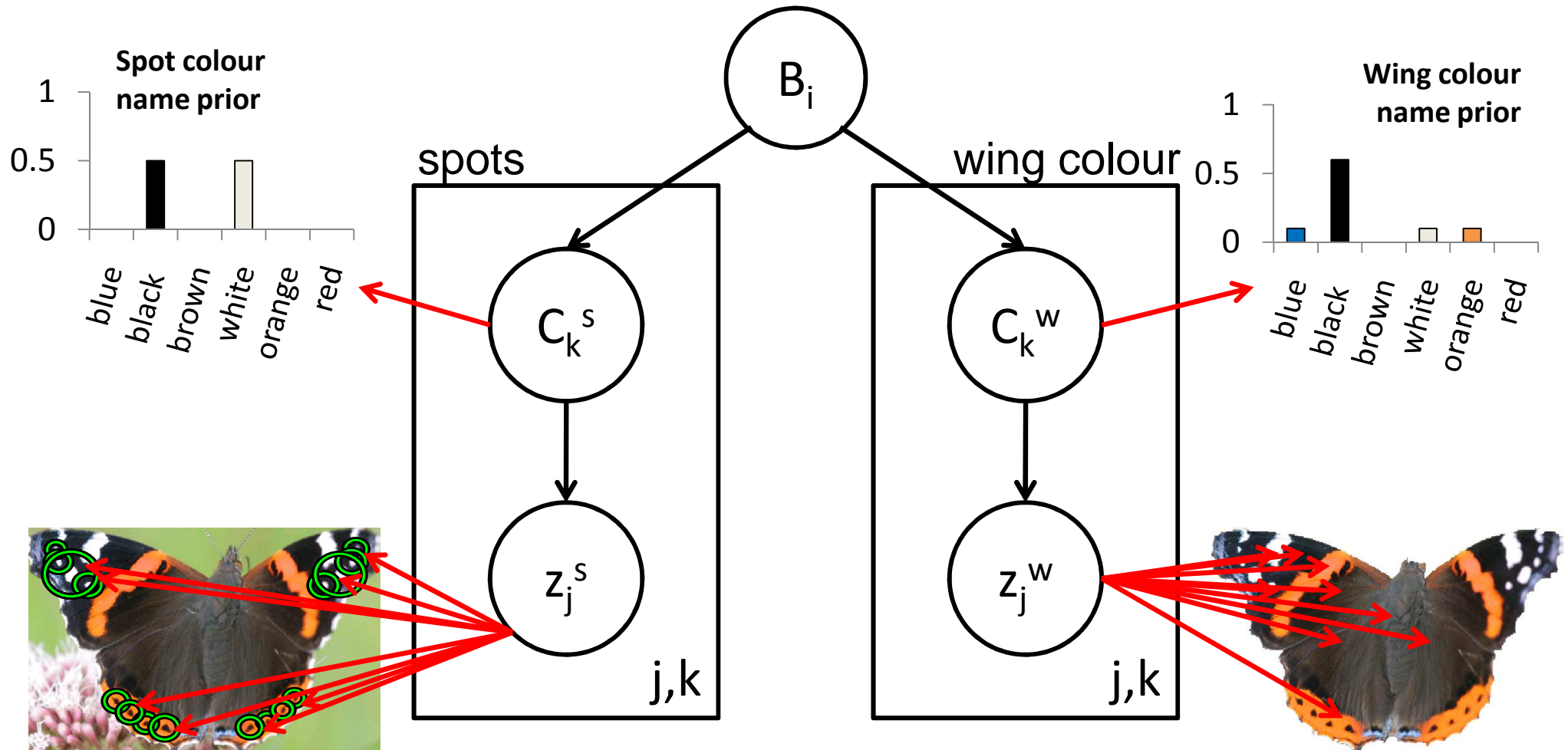
+ (1 - α)

'Other' Colour



Generative Model

$$p(I/B_i) = p(S/B_i) p(W/B_i)$$



Classification: Assign to category which maximises $p(I/B_i)$

Experimental Results

Humans

As “Upper Bound”

Butterfly Recognition Experiment

Please read the description below and select the butterfly that matches the description by clicking on the corresponding image. There is one (and only one) correct image. You can change your answer by clicking on another image. **Please read the description carefully - you have only one chance!**

When you have selected an image, please complete the information at the bottom of the page - tell us if you are a native English speaker and/or an expert on butterflies, enter your email address, and click Submit.

Description

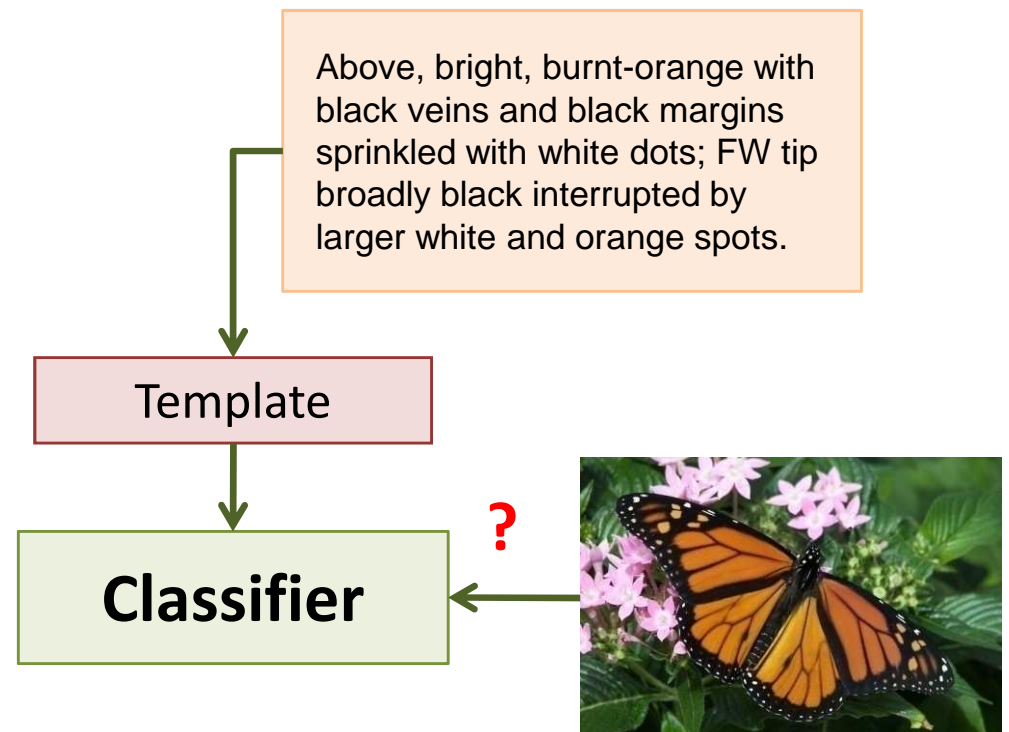
Wings long and narrow. Jet-black above, banded with lemon-yellow (sometimes pale yellow). Beneath similar; bases of wings have crimson spots.



I am a native English speaker
 I am an expert on butterflies
Email address:

Proposed Method

How well can machine learn from only textual descriptions?



Human Performance

Butterfly Recognition Experiment

Please read the description below and select the butterfly that matches the description by clicking on the corresponding image. There is one (and only one) correct image. You can change your answer by clicking on another image. **Please read the description carefully - you have only one chance!**

When you have selected an image, please complete the information at the bottom of the page - tell us if you are a native English speaker and/or an expert on butterflies, enter your email address, and click Submit.

Description

Wings long and narrow. Jet-black above, banded with lemon-yellow (sometimes pale yellow). Beneath similar; bases of wings have crimson spots.



I am a native English speaker

I am an expert on butterflies

Email address:

Limited to **ONE** single trial

Human Performance

Predicted

	Predicted										
	danaus plexippus	heliconius charitonius	heliconius erato	junonia coenia	lycaena phlaeas	nymphalis antiopa	papilio cresphontes	peris rapae	vanessa atalanta	vanessa cardui	
danaus plexippus	87									13	23
heliconius charitonius	13	31	19				38				16
heliconius erato			96							4	23
junonia coenia				86	5					9	22
lycaena phlaeas				24	76						21
nymphalis antiopa	4		7	4	81			4			27
papilio cresphontes	14		7			71				7	14
peris rapae	6						94				16
vanessa atalanta	15			20	5	5		55			20
vanessa cardui	16			5	5	11			21	42	19

Native speakers: 72%
(201 participants)

Predicted

	Predicted										
	danaus plexippus	heliconius charitonius	heliconius erato	junonia coenia	lycaena phlaeas	nymphalis antiopa	papilio cresphontes	peris rapae	vanessa atalanta	vanessa cardui	
danaus plexippus	67									33	3
heliconius charitonius	33	17	17							17	6
heliconius erato	25		38	13	13				13		8
junonia coenia				75					25		8
lycaena phlaeas					88	13					8
nymphalis antiopa	25			25	50						4
papilio cresphontes	50					50					4
peris rapae	20		20				60				5
vanessa atalanta				33				33	33		3
vanessa cardui	33					33				33	3

Non-native speakers: 51%
(52 participants)

Chance performance: 10%

Example Misclassification

Heliconius charitonius (Zebra Longwing)

Wings long and narrow. Jet-black above, banded with lemon-yellow (sometimes pale yellow). Beneath similar; bases of wings have crimson spots.



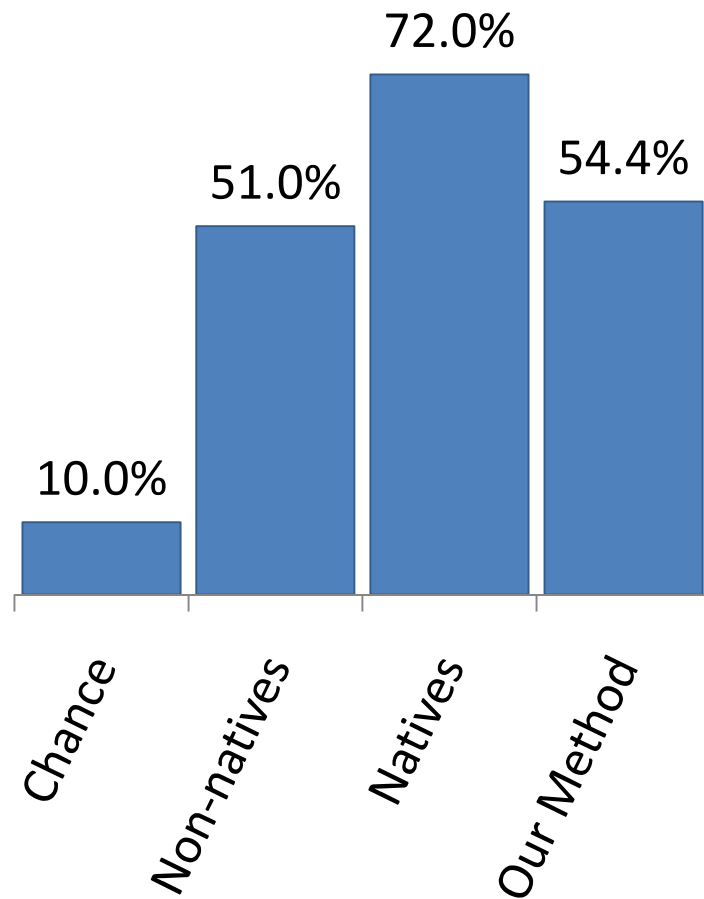
spots are not mentioned
'lemon-yellow' bands?

Confused with →

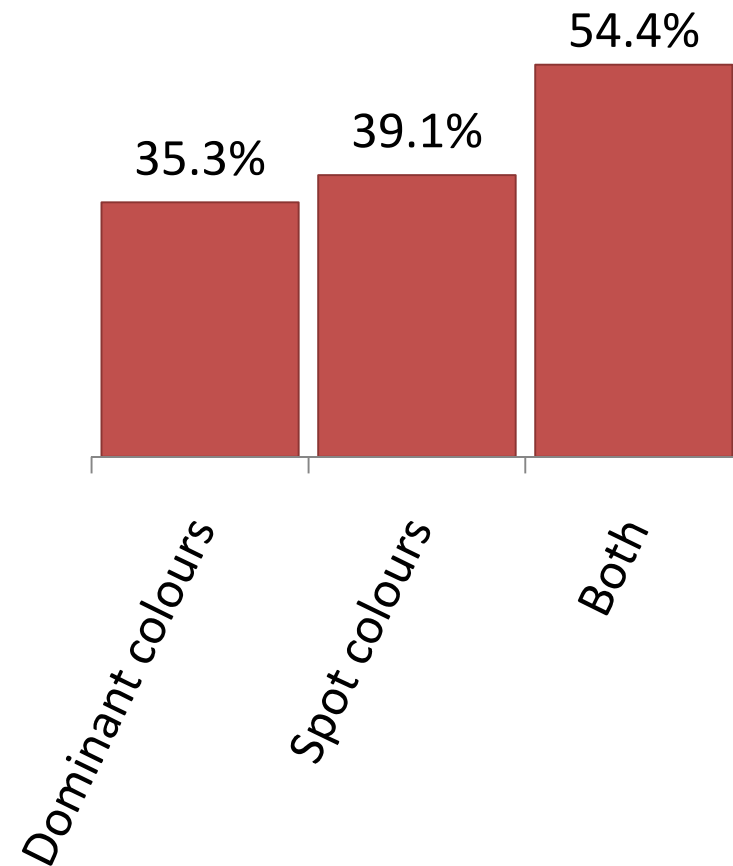


Results: Proposed Method

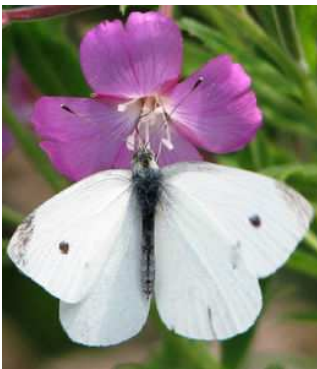
Humans vs. Our Method



Individual vs. Combined Components



Results: Proposed Method



Predicted

danaus plexippus
heliconius charitoni
heliconius erato
junonia coenia
lycaena phlaeas
nymphalis antiopa
papilio cresphontes
peris rapae
vanessa atalanta
vanessa cardui

Ground truth	<i>danaus plexippus</i>	<i>heliconius charitoni</i>	<i>heliconius erato</i>	<i>junonia coenia</i>	<i>lycaena phlaeas</i>	<i>nymphalis antiopa</i>	<i>papilio cresphontes</i>	<i>peris rapae</i>	<i>vanessa atalanta</i>	<i>vanessa cardui</i>	
<i>danaus plexippus</i>	35				4	18	41	1			82
<i>heliconius charitoni</i>		0	1		1	42	55	1			93
<i>heliconius erato</i>			49	10	7	2	33				61
<i>junonia coenia</i>	1			2	42	20	12	2	20		90
<i>lycaena phlaeas</i>	3				94	2					88
<i>nymphalis antiopa</i>			1		1	61	2	3	12	20	100
<i>papilio cresphontes</i>							85	15			89
<i>peris rapae</i>								93			55
<i>vanessa atalanta</i>	2								84	13	90
<i>vanessa cardui</i>	6				15				39	39	84

Accuracy: 54.4%

Results: Proposed Method



Predicted

	Predicted										
	<i>danaus plexippus</i>	<i>heliconius charitoni</i>	<i>heliconius erato</i>	<i>junonia coenia</i>	<i>lycaena phlaeas</i>	<i>nymphalis antiopa</i>	<i>papilio cressphontes</i>	<i>peris rapae</i>	<i>vanessa atalanta</i>	<i>vanessa cardui</i>	
<i>danaus plexippus</i>	35	0	0	0	0	0	0	0	0	0	82
<i>heliconius charitoni</i>	0	1	0	0	0	0	0	0	0	0	93
<i>heliconius erato</i>	0	0	49	0	0	0	0	0	0	0	61
<i>junonia coenia</i>	1	0	0	2	0	0	0	0	0	0	90
<i>lycaena phlaeas</i>	3	0	0	0	94	0	0	0	0	0	88
<i>nymphalis antiopa</i>	0	0	1	0	0	61	0	0	0	0	100
<i>papilio cressphontes</i>	0	0	0	0	0	0	85	0	0	0	89
<i>peris rapae</i>	0	0	0	0	0	0	0	7	0	0	55
<i>vanessa atalanta</i>	0	0	0	0	0	0	0	0	84	0	90
<i>vanessa cardui</i>	0	0	0	0	0	0	0	0	0	39	84

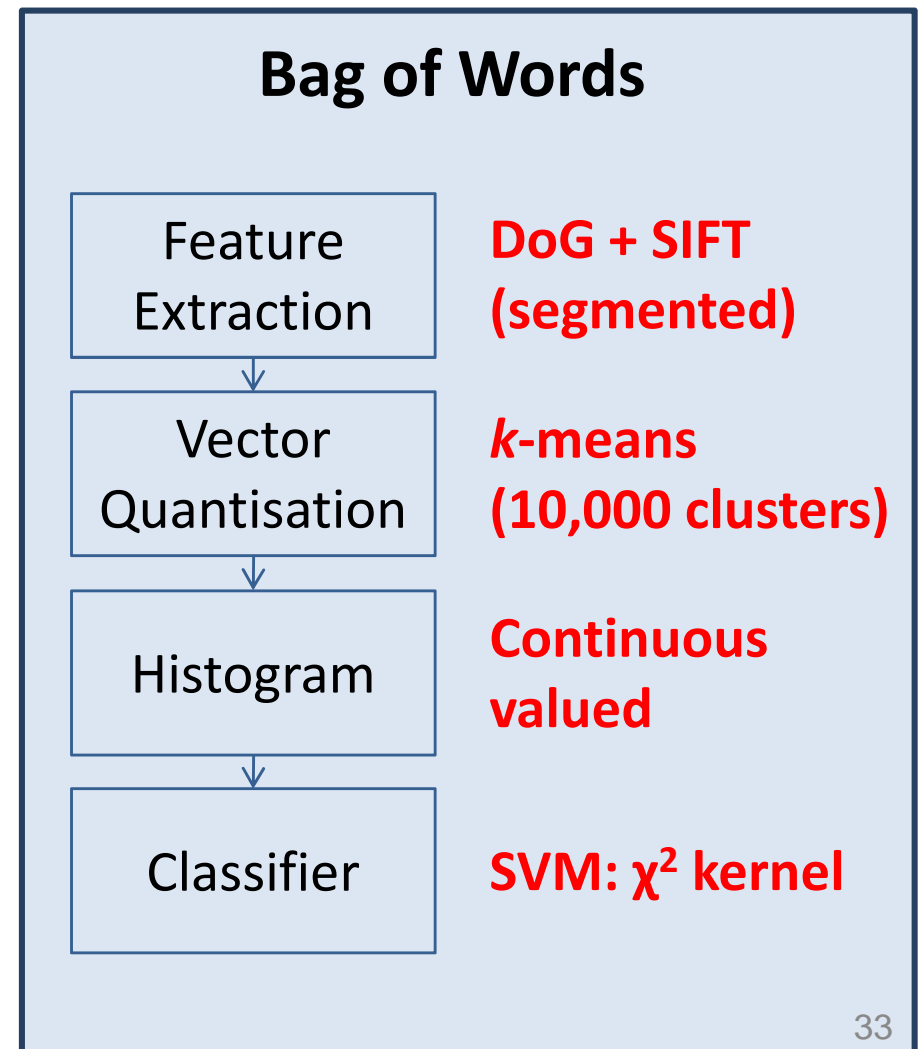
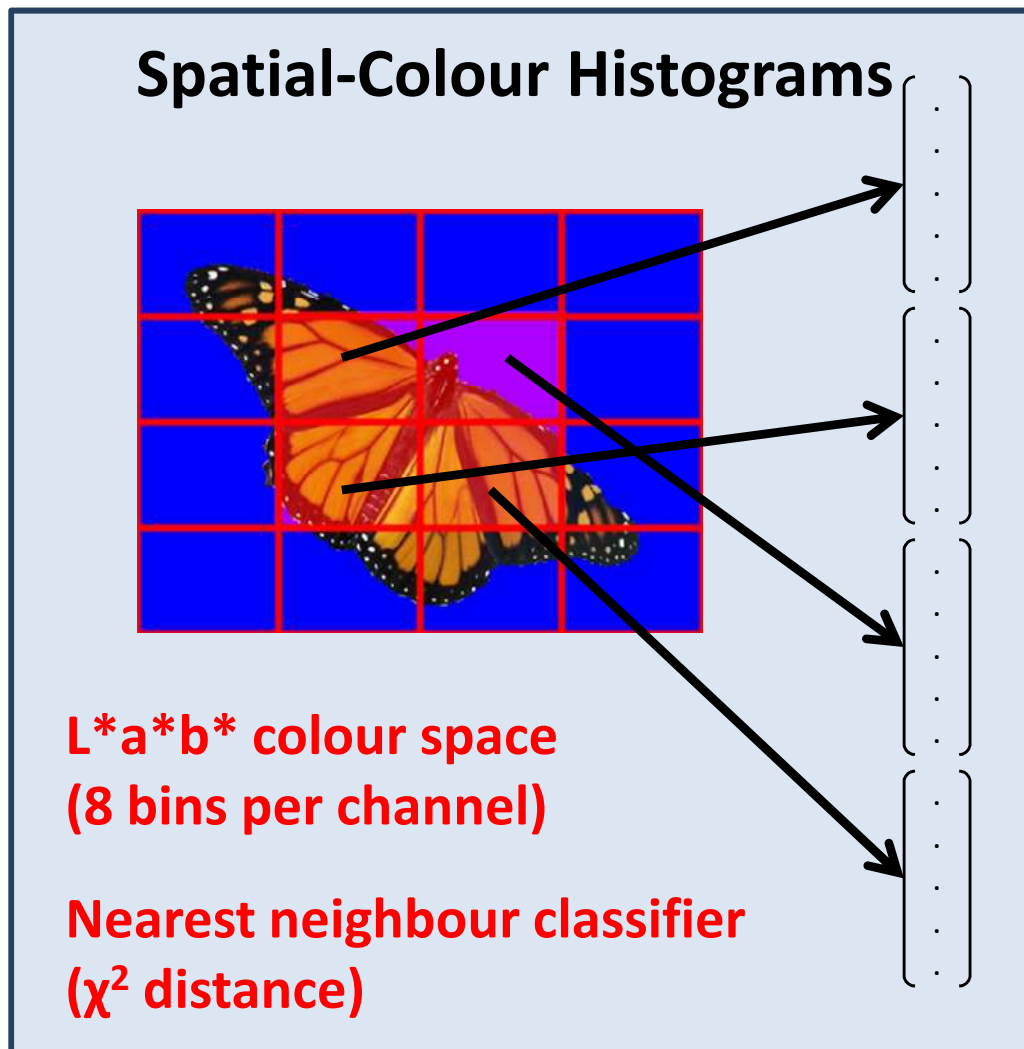
Ground truth

Accuracy: 54.4%

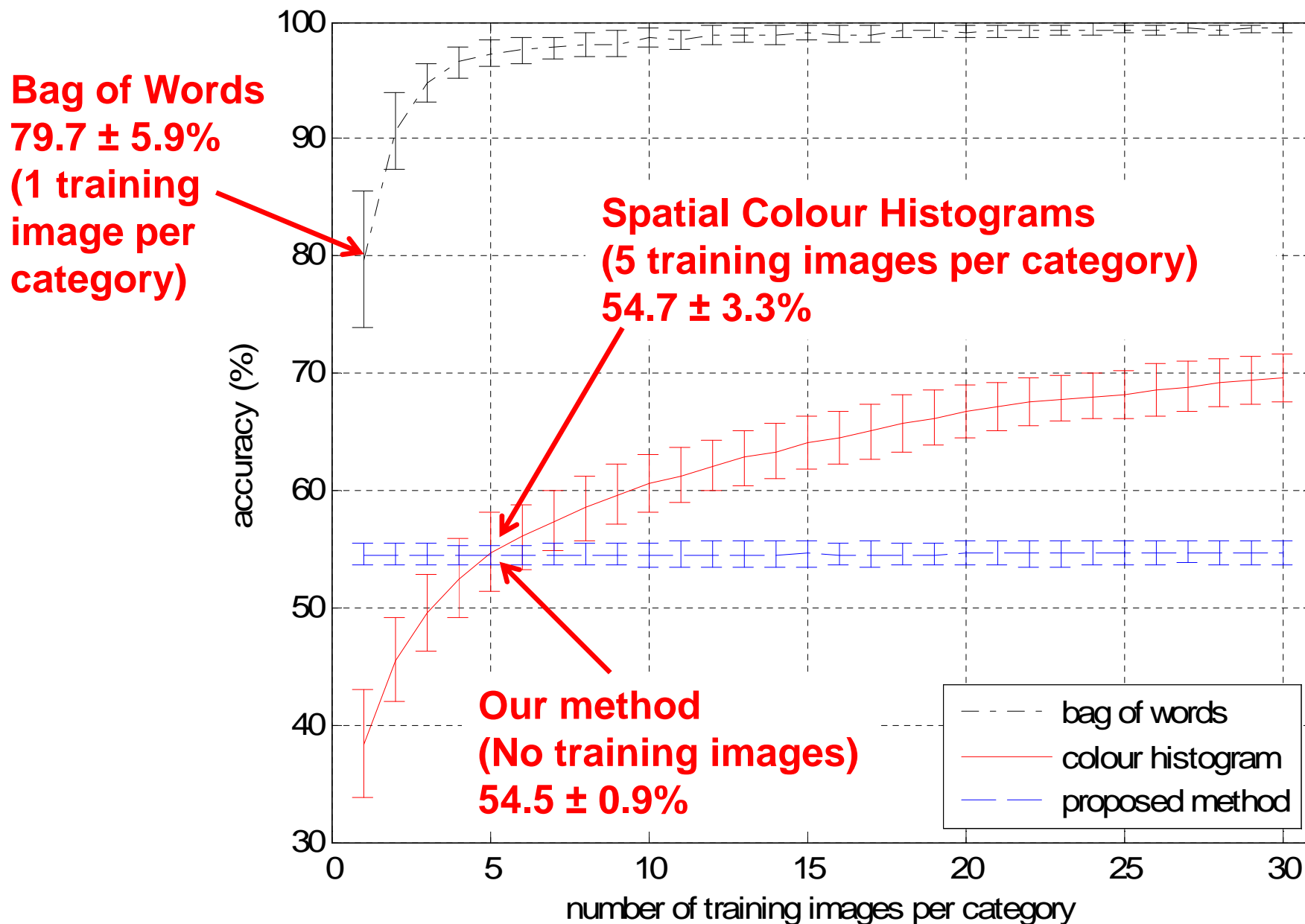


Standard Vision Methods

- Proposed method is compared against two standard approaches:



Results: Standard Vision Methods



Discussion

- We investigated models for linking information in text and images together
- Mapping between textual and image features is a challenging problem
- Initial model achieved modest accuracy with no training images
- State of the art vision methods give good results but depend on the training images used
- Future work:
 - Extract more information from text
 - Combine information from multiple texts
 - **Combine text with images**



Learning Models for Object Recognition from Natural Language Descriptions

Josiah Wang
Katja Markert
Mark Everingham

School of Computing
University of Leeds