

Descriptive Attributes for Language-based Object Keypoint Detection



Jerod Weinman



Serge Belongie



Stella Frank

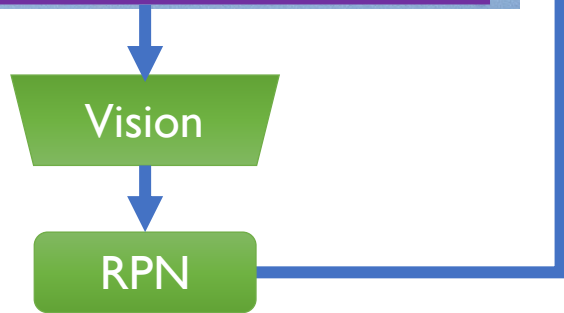


Grinnell College
(USA)

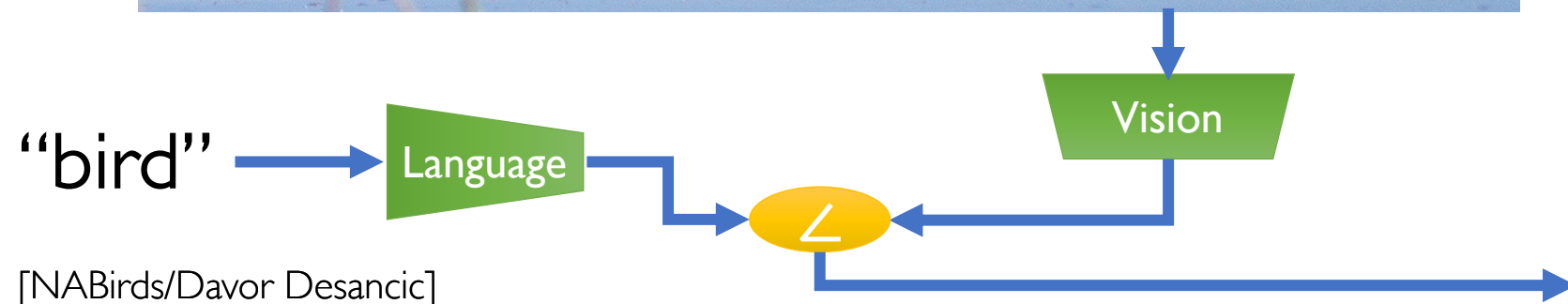
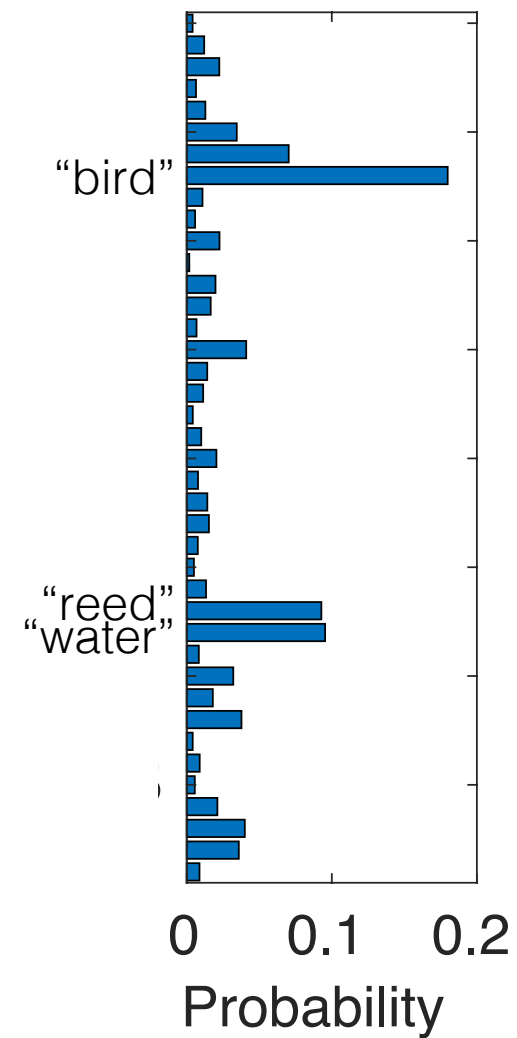


University of Copenhagen / Pioneer Centre for AI
(Denmark)





R-CNN
2014

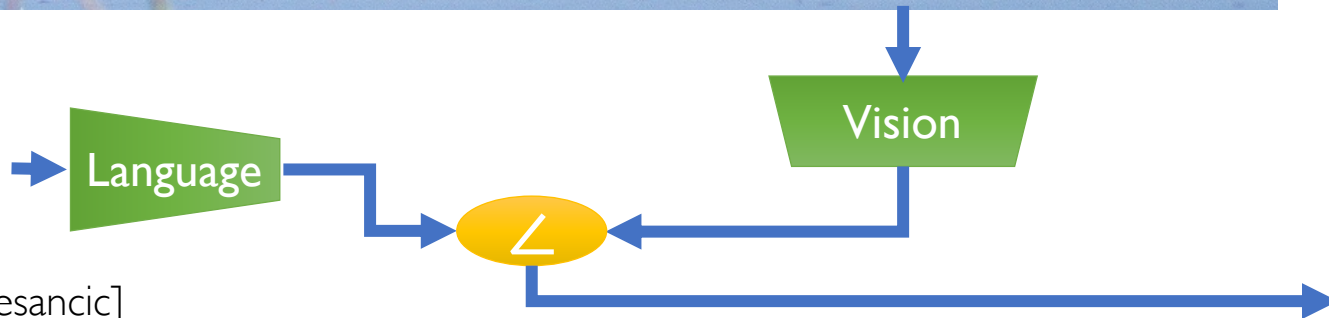


CLIP
2021

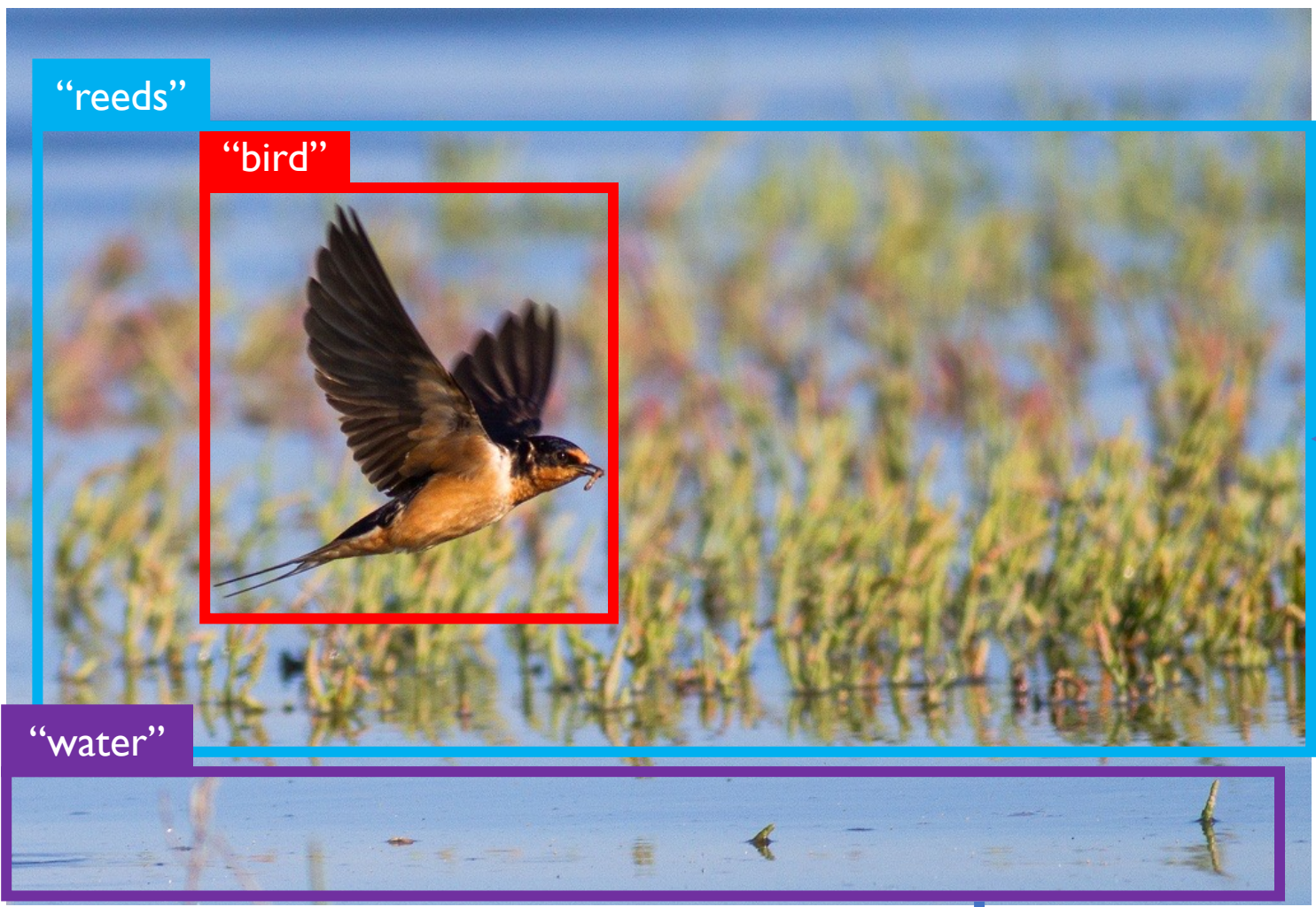


“barn
swallow”

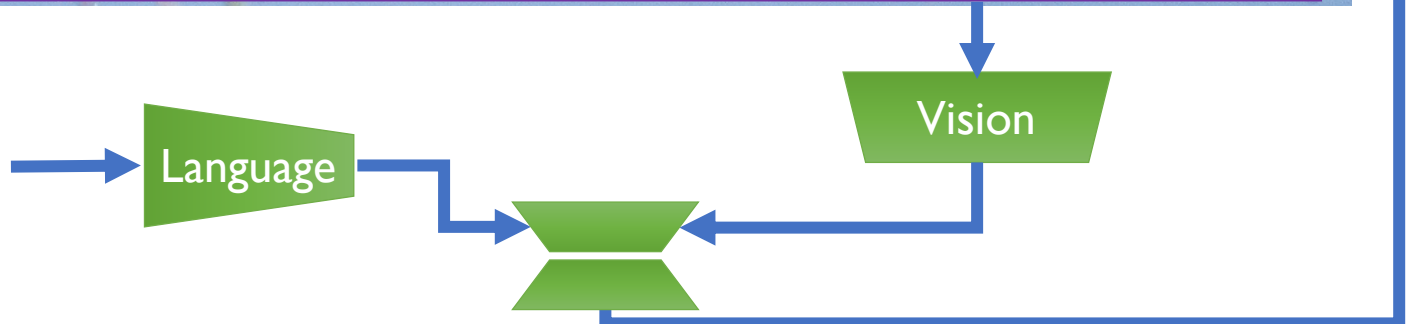
[NABirds/Davor Desancic]



CLIP
2021



“bird.
reeds.
water.”



MDETR
2021

GLIP
2022



“flying bird.
reeds.
water.”

Language

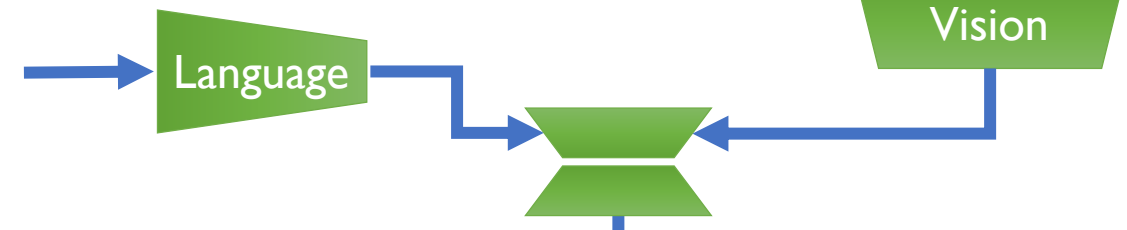
Vision

MDETR
2021

GLIP
2022



“flying barn swallow.
reeds.
water.”

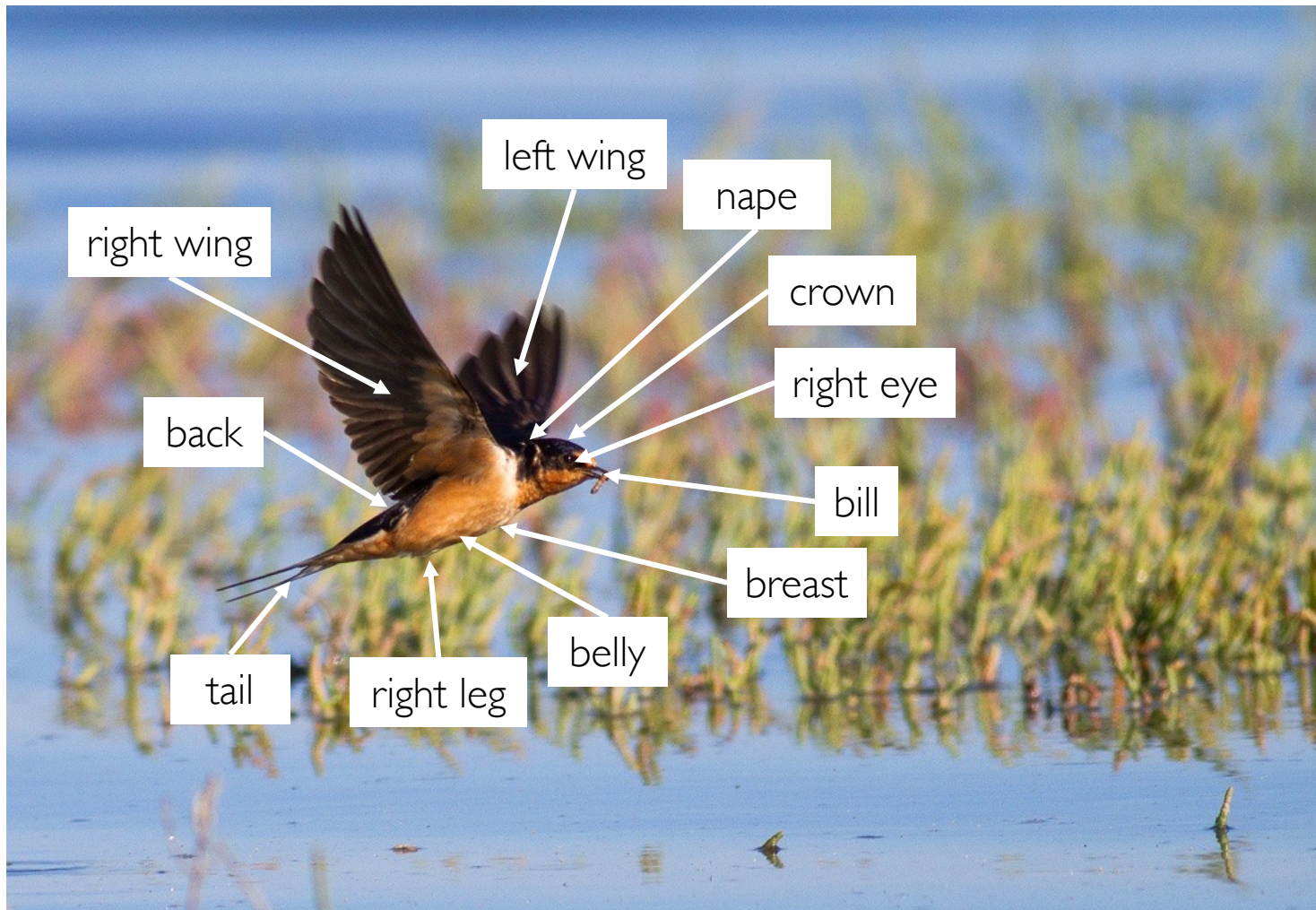


MDETR
2021

GLIP
2022

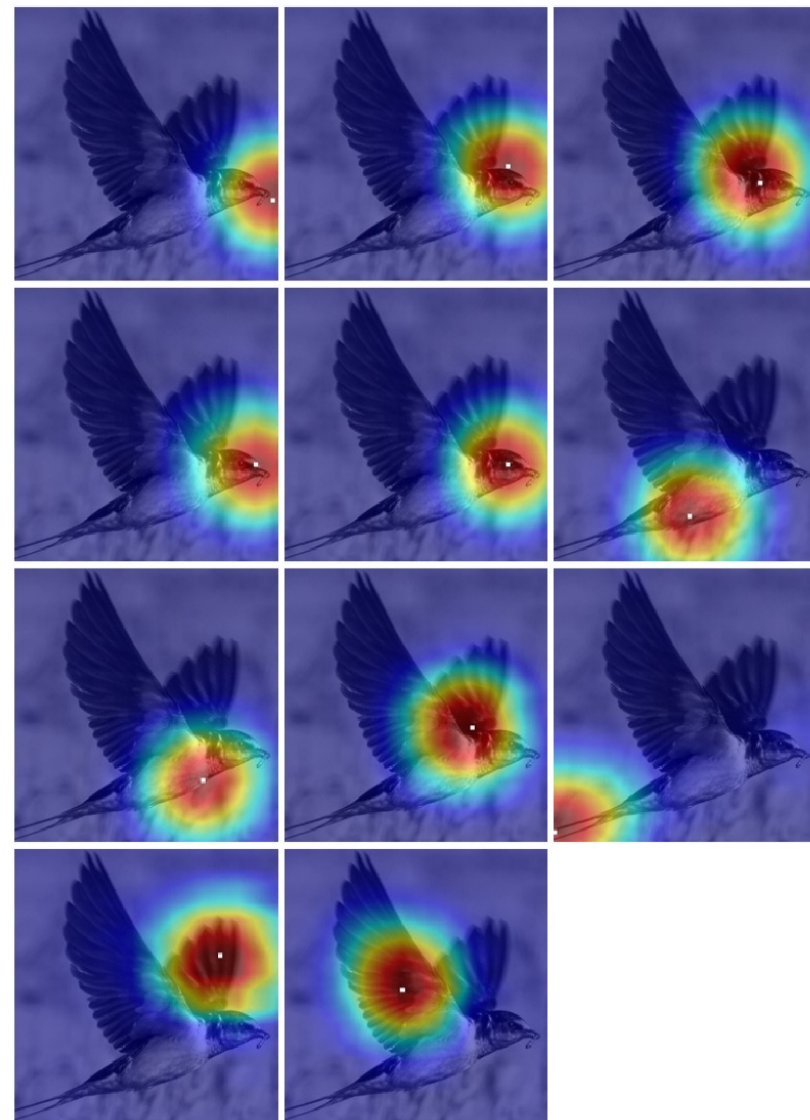


Keypoint Detection

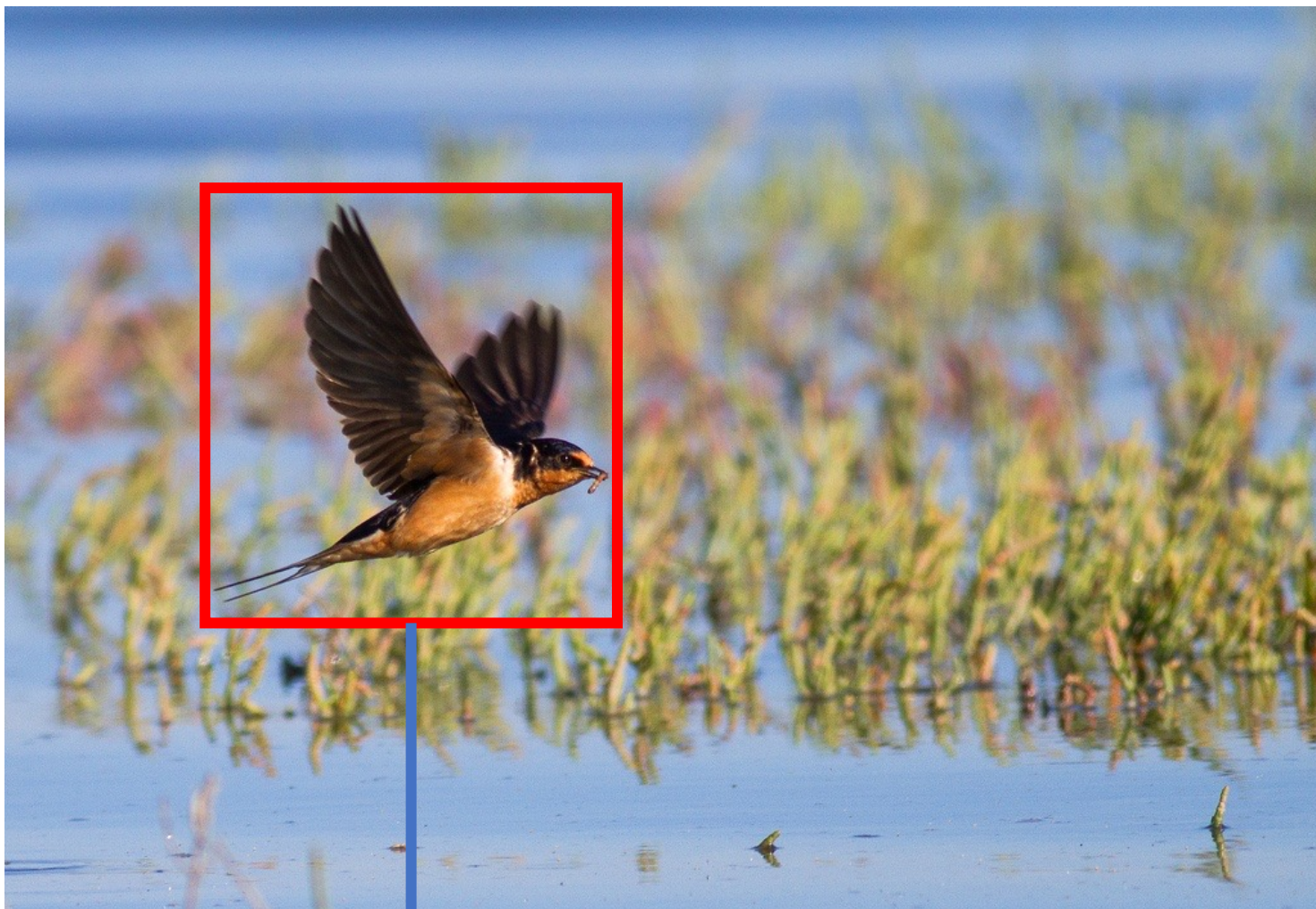




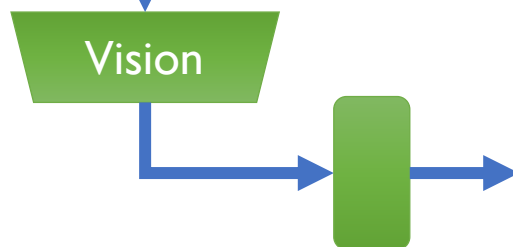
Vision



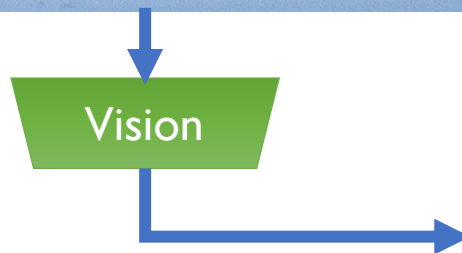
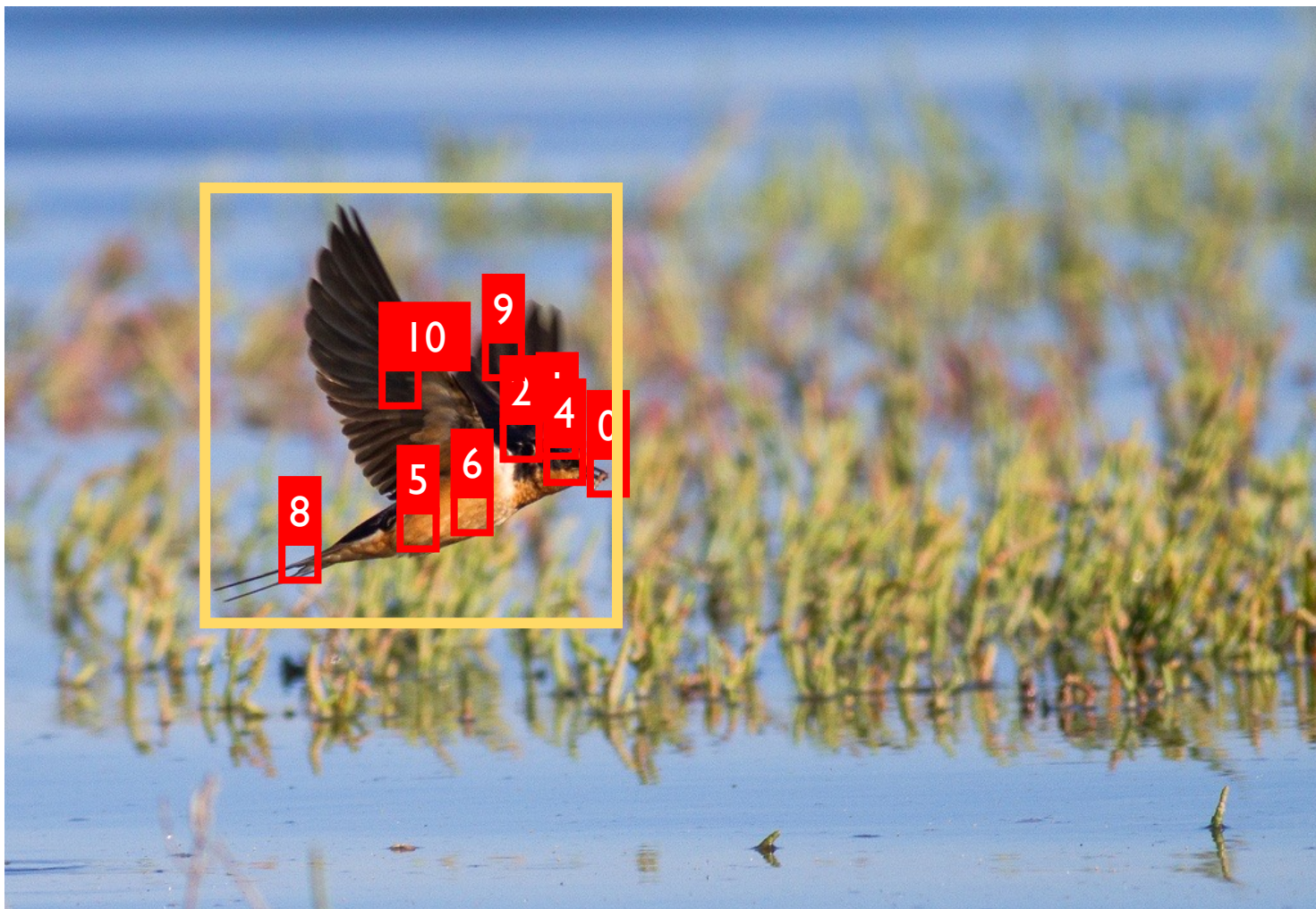
Heatmaps
2014



0: 465 364
1: 432 340
2: 402 341
3: - -
4: 433 354
5: 344 413
6: 385 398
7: 389 343
8: 239 435
9: 404 271
10: 314 307



DeepPose
2014



Keypoints And Poses As Objects

KAPAO
2022



“bird.
reeds.
water.”



KP-GLIP

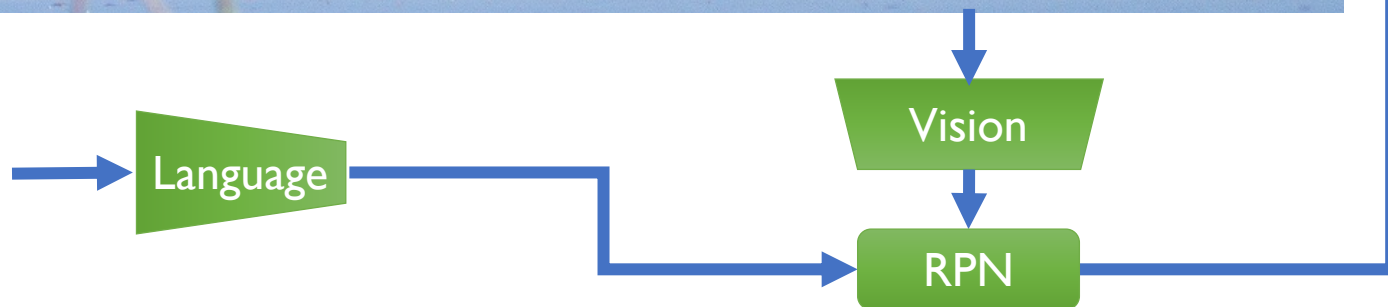


RQ I: How well does a VL model designed for object detection perform at keypoint detection?

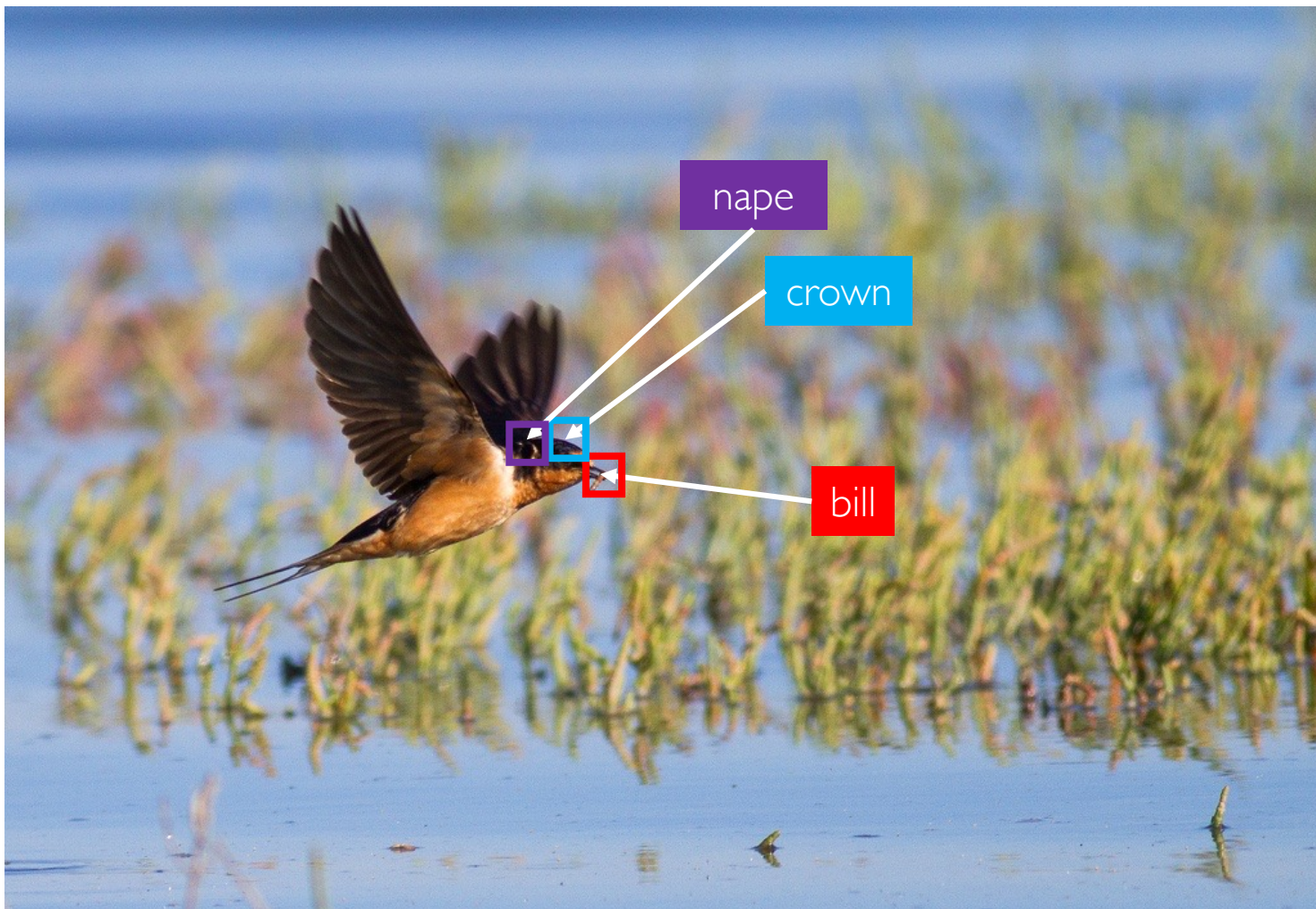
NABirds: (2015)

- 50K images
- 400 species
- 11 keypoints

“bird.
reeds.
water.”



KP-GLIP



RQ I: How well does a VL model designed for object detection perform at keypoint detection?

NABirds: (2015)

- 50K images
- 400 species
- 11 keypoints

KP-GLIP

“bill.
crown.
nape....”

Language

Vision

RPN





RQ I: How well does a VL model designed for object detection perform at keypoint detection?












NABirds:

- 50K images
- 400 species
- 11 keypoints

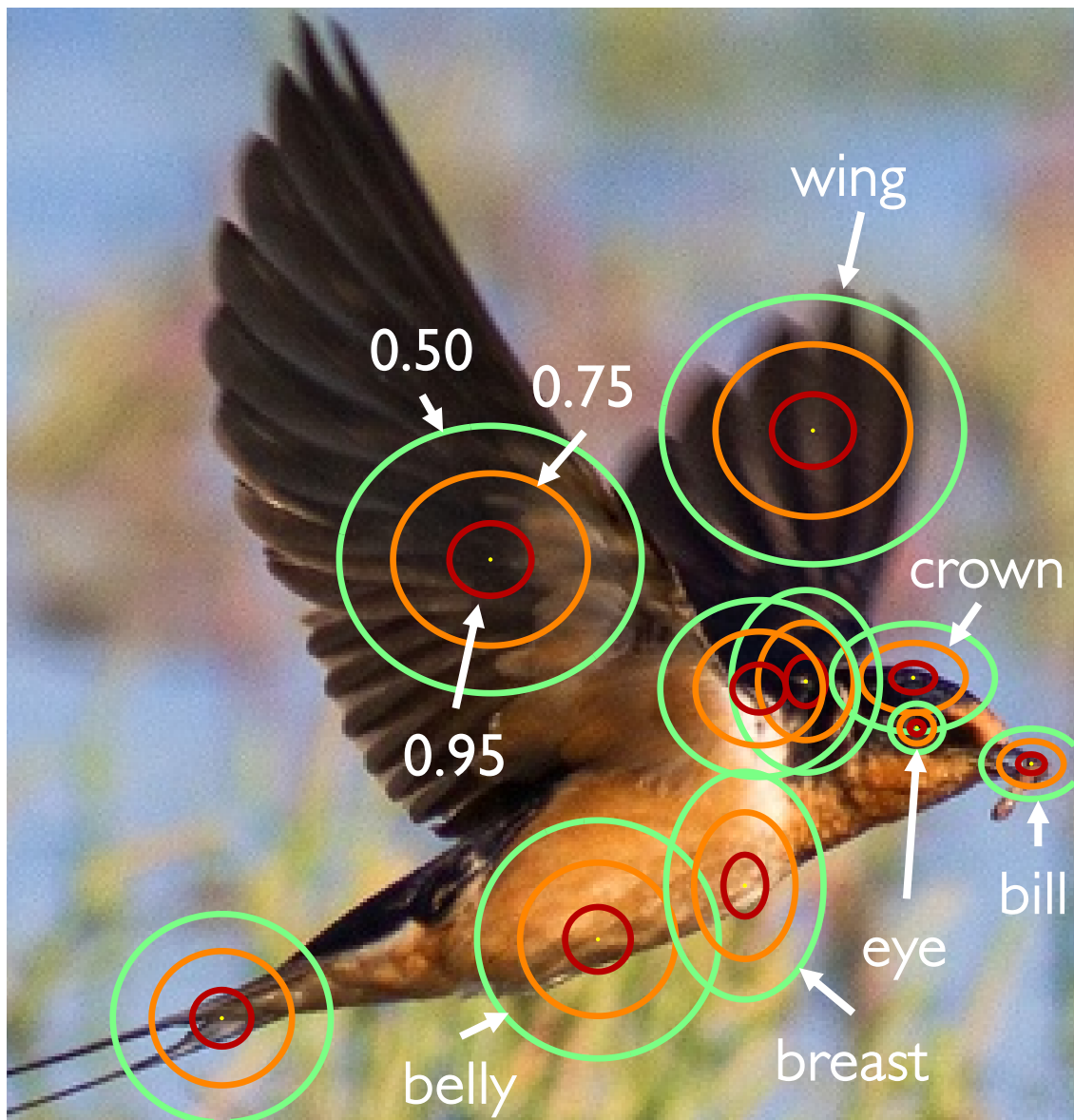
Names: bill. crown. nape. left eye. right eye. belly. breast. back. tail. left wing. right wing.

Symbols: f. g. h. k. m. n. p. q. r. w. y.

Evaluation

Metric	False Positives	Dist.Threshold	
PCK	 Ignored	 Fixed	
COCO IOU	 Precision	 Sweeps	 for boxes
COCO OKS	 Ignored	 Sweeps	 annotator variance
OKS mAP	 Precision	 Sweeps	 anisotropic variance

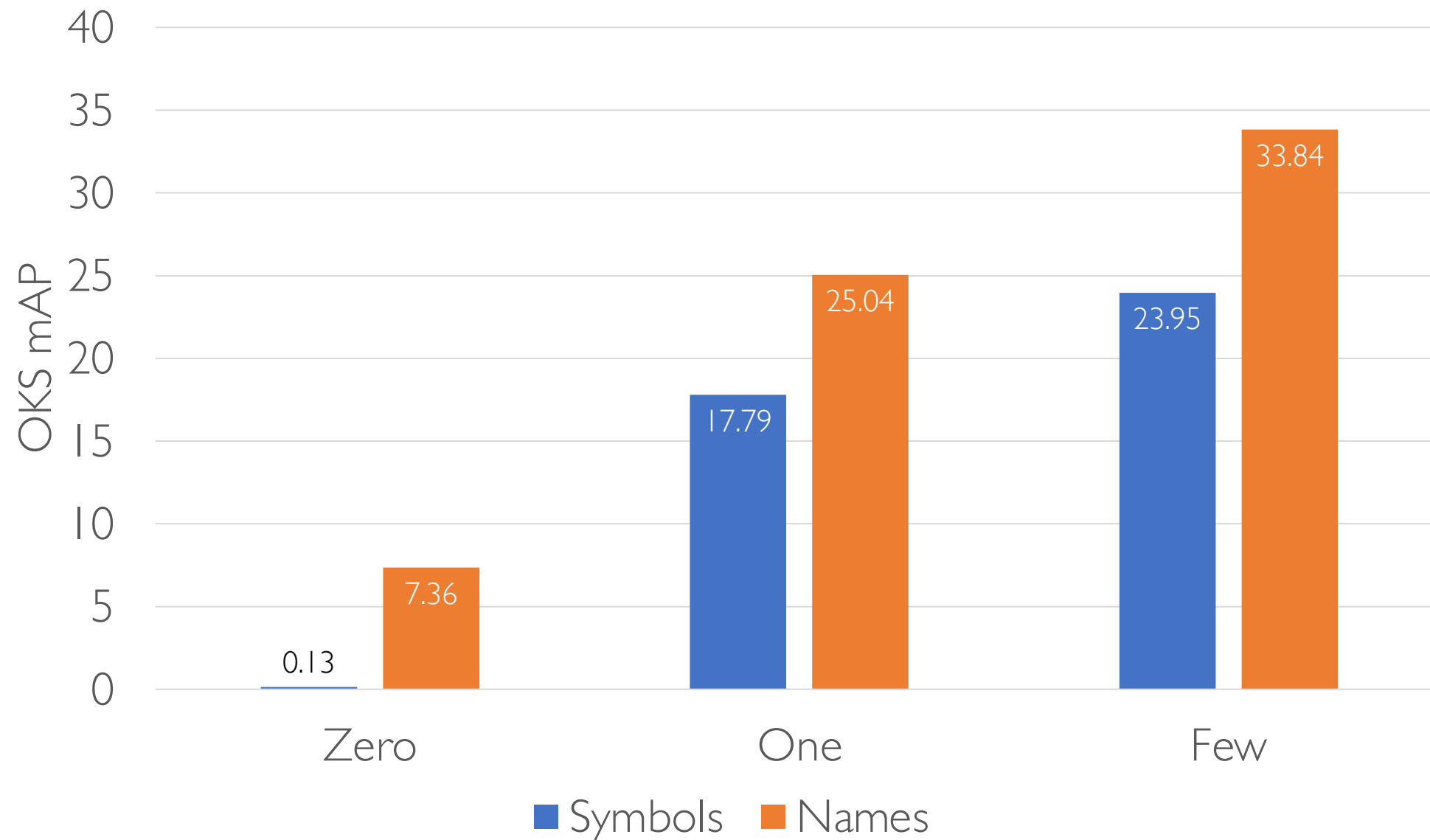
Evaluation



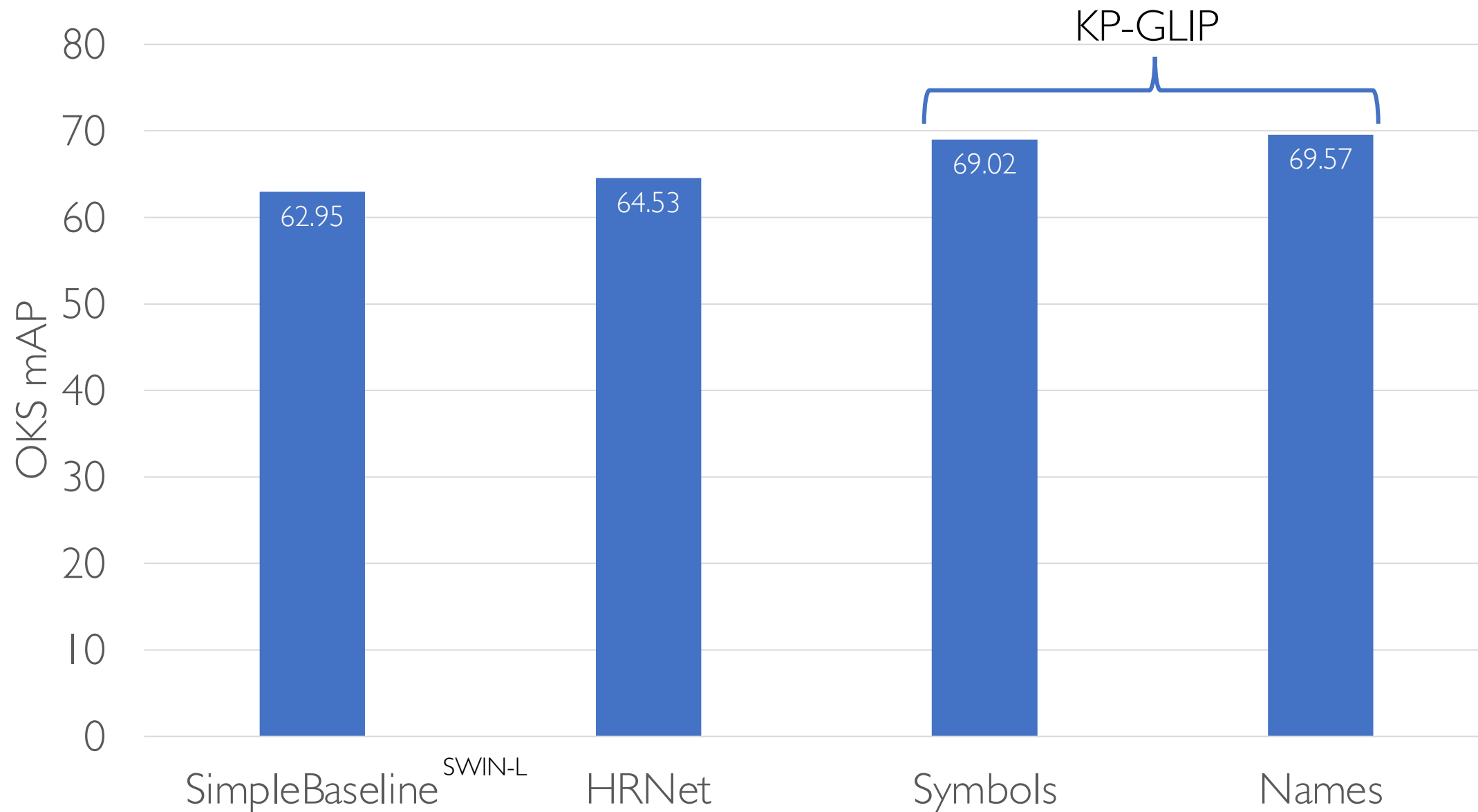
OKS mAP

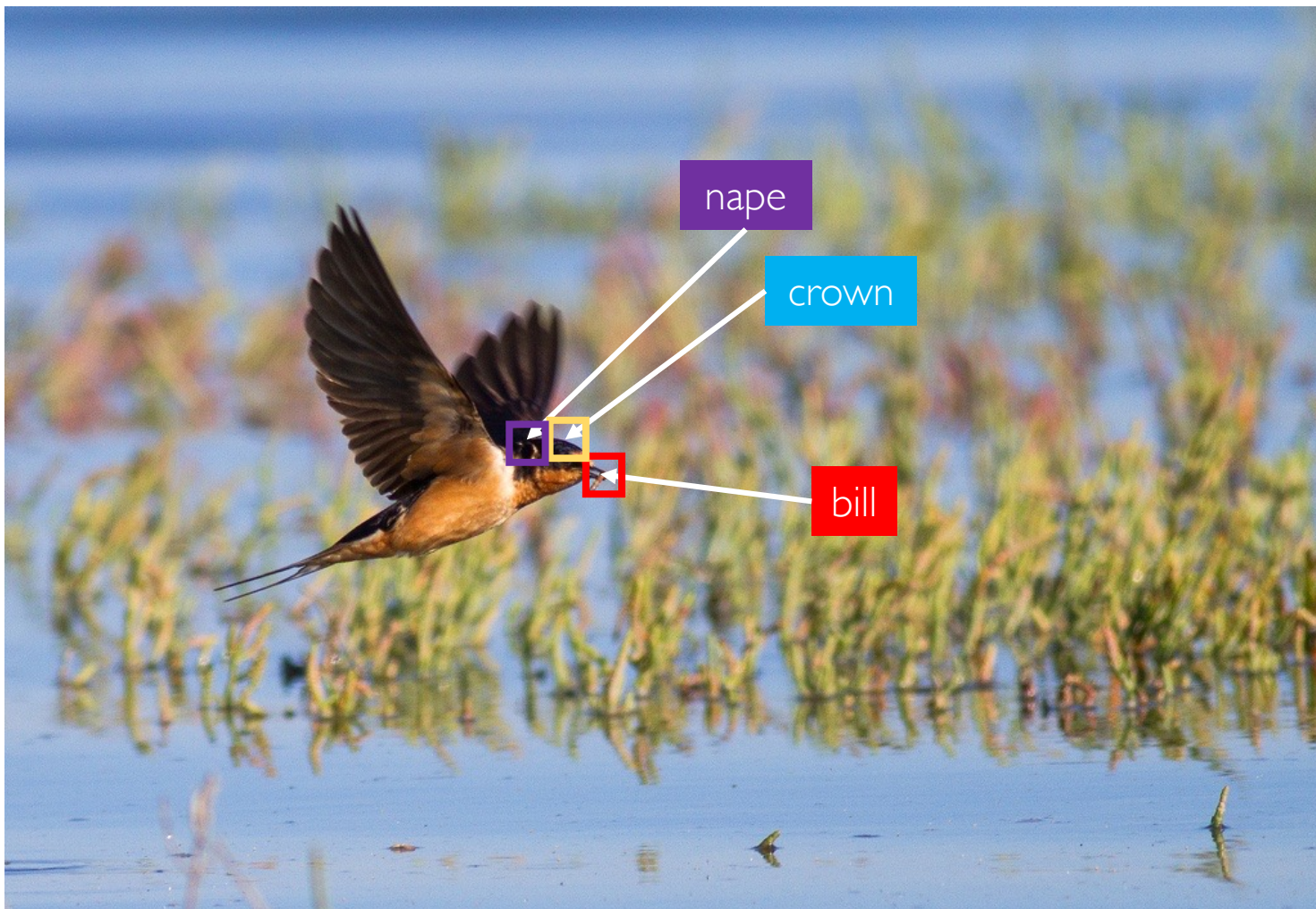
- ✓ Precision penalizes false positives
- ✓ mAP sweeps distance thresholds
- ✓ Anisotropic distance threshold varies with annotator variation

Finetuned N-Shot Learning



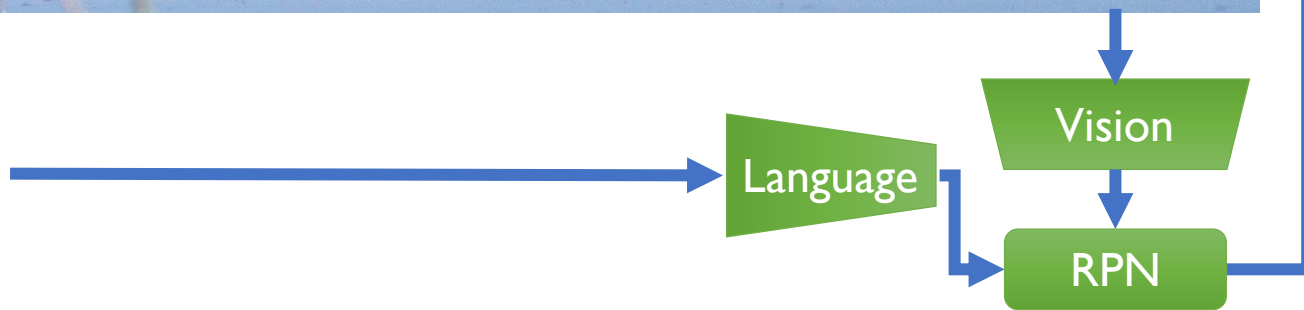
Finetuning Results





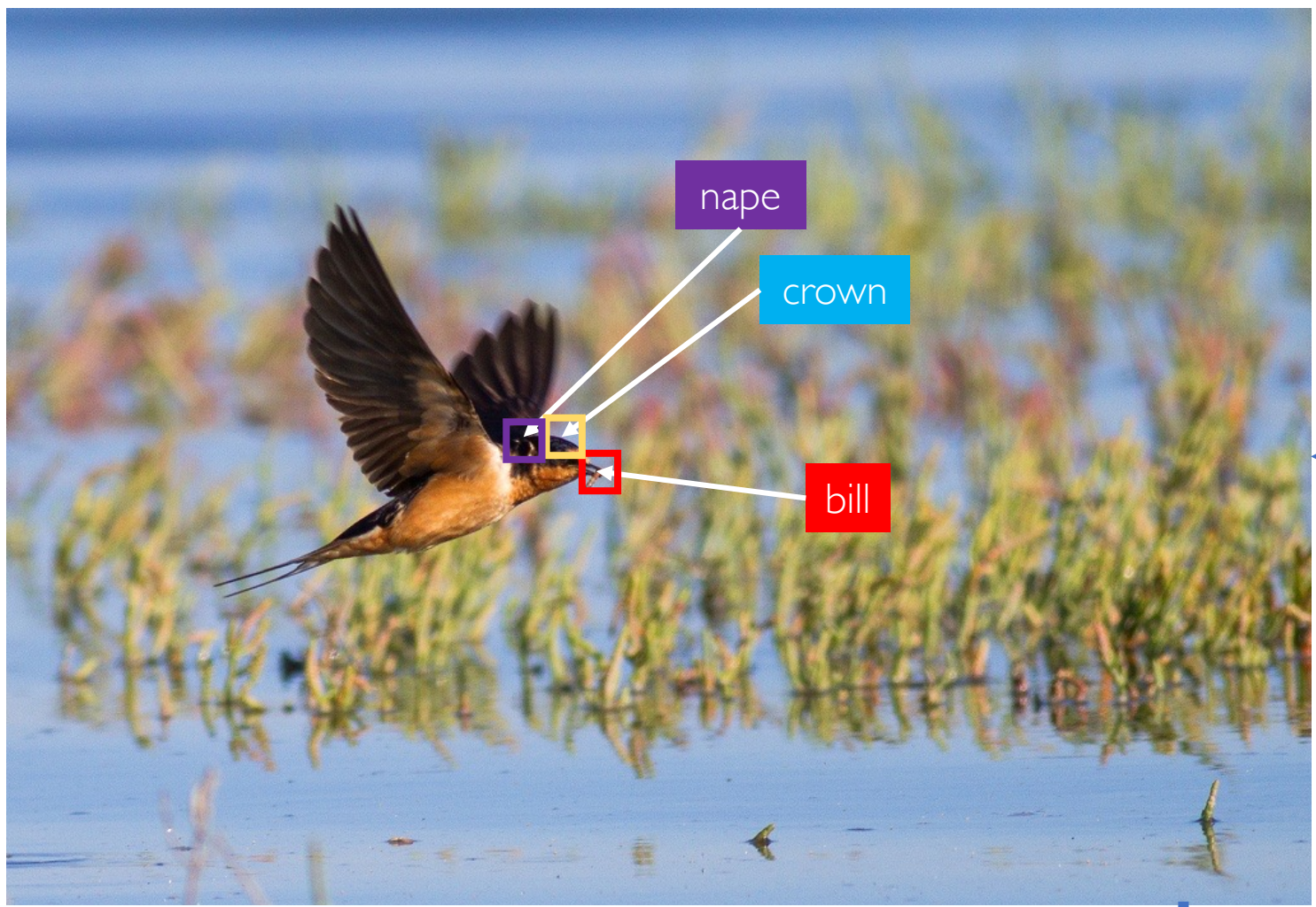
RQI: How well does a VL model designed for object detection perform at keypoint detection?

“bill.
crown.
nape....”

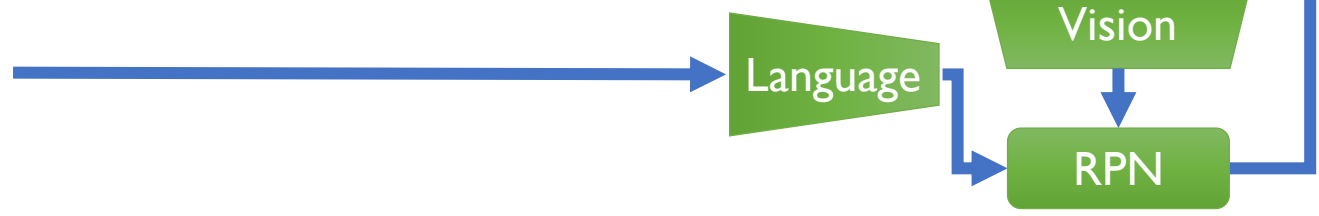


KP-GLIP

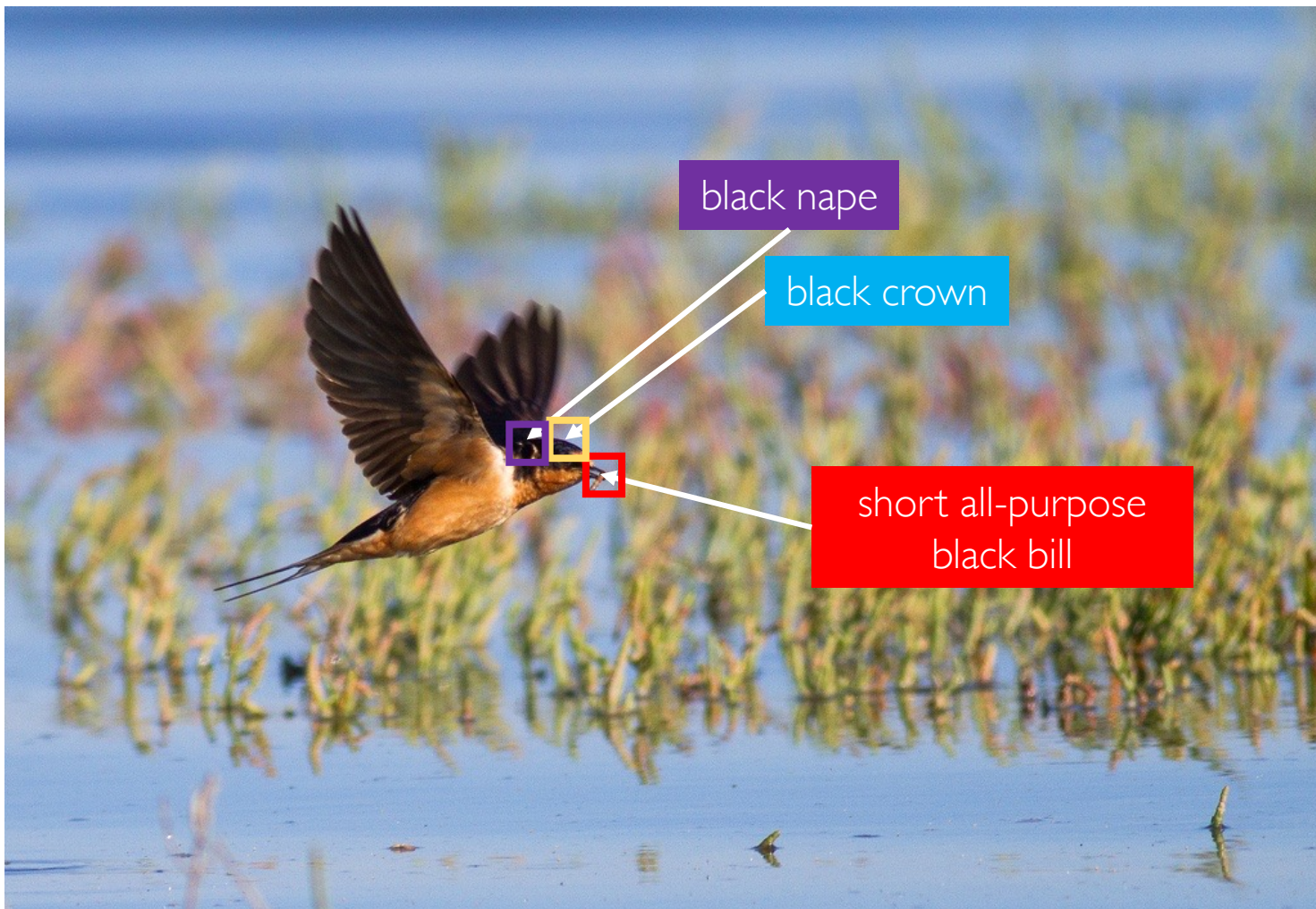
RQ2: Does adding descriptive attributes to keypoint names improve performance?



“bill.
crown.
nape....”

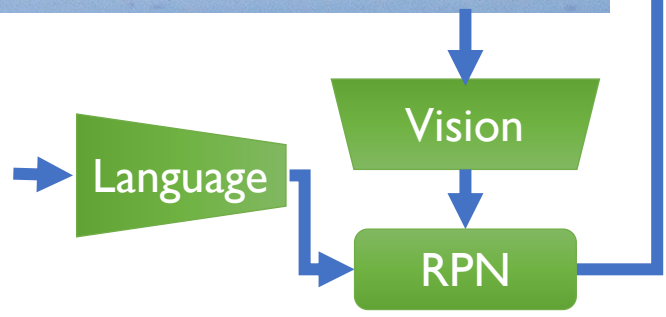


KP-GLIP



RQ2: Does adding descriptive attributes to keypoint names improve performance?

“short all-purpose black bill.
black crown.
black nape....”

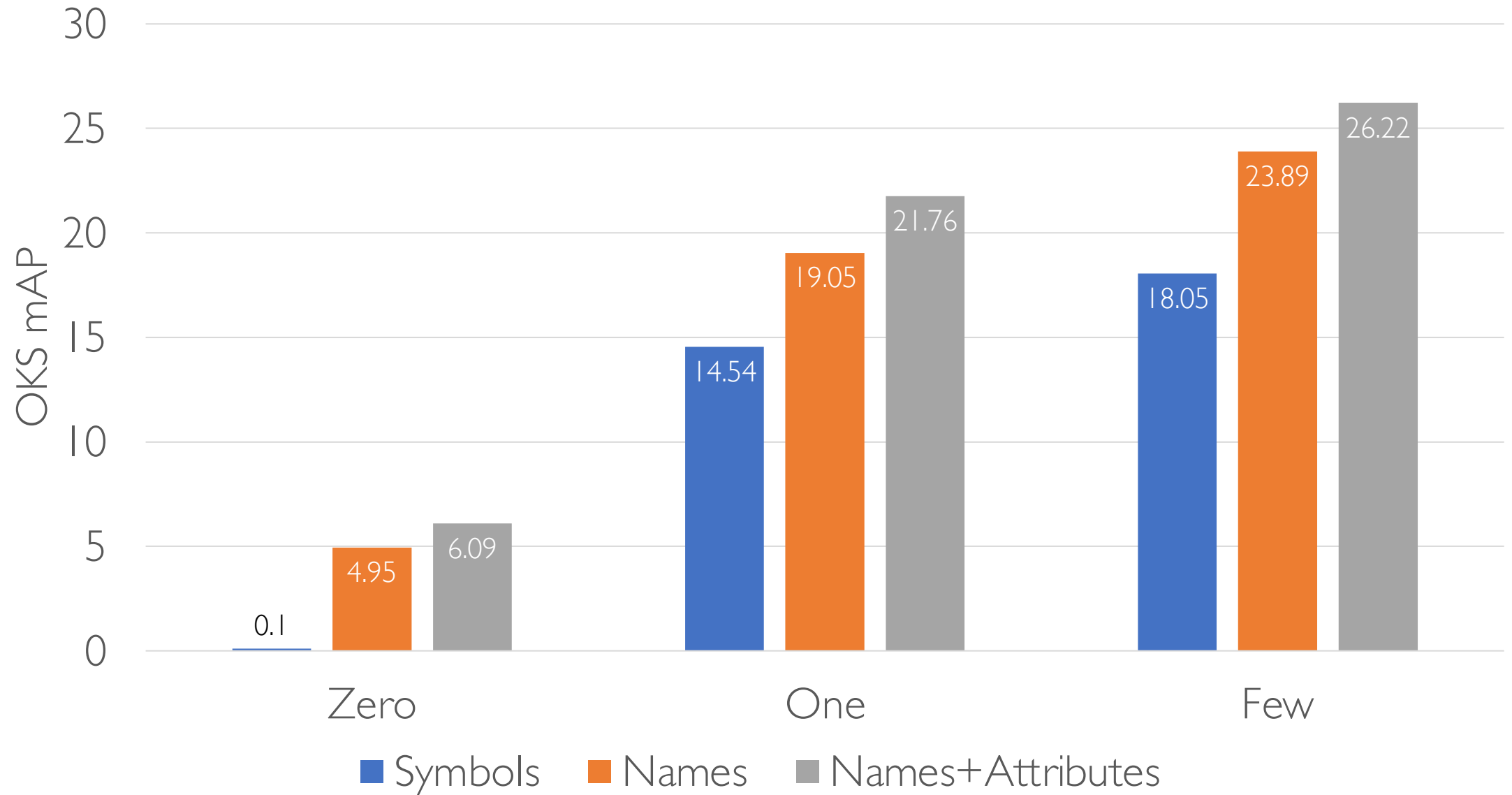


KP-GLIP

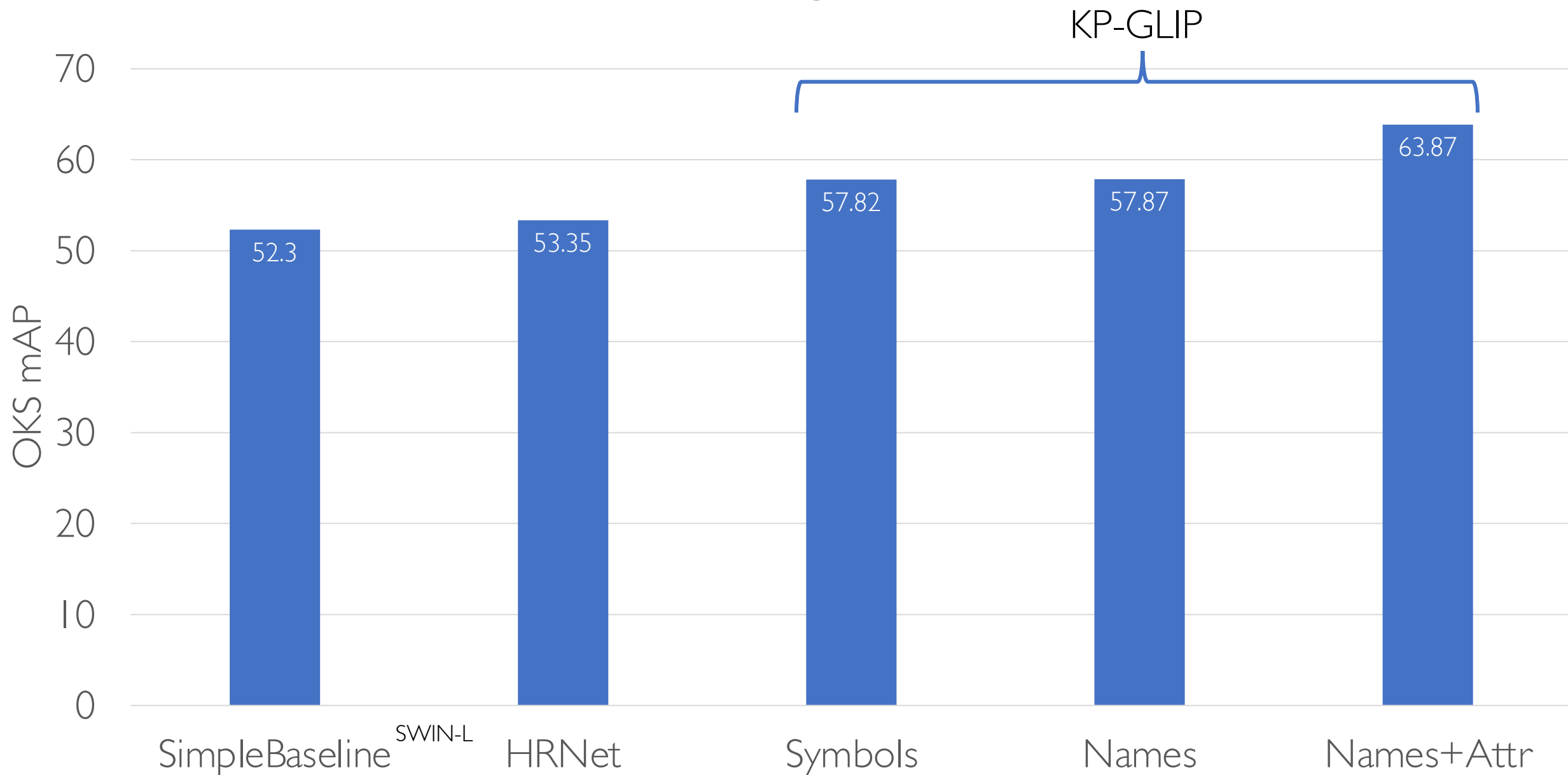
CUB Descriptive Keypoint Attributes

Name	Attributes	Examples
bill	length / shape / color(s)	short all-purpose grey bill. long needle black bill.
crown	color(s)	black and white crown. blue crown.
nape	color(s)	black and white nape. brown and black nape. buff nape.
eye	color(s)	black and red left eye. black right eye. yellow right eye.
belly	pattern / color(s)	striped brown and white belly. solid belly.
breast	pattern / color(s)	striped yellow and black breast. white breast.
back	pattern / color(s)	striped brown black and buff back. solid blue back.
tail	pattern / shape / color(s)	solid notched tail. notched brown tail.
wing	pattern / shape / color(s)	spotted pointed black and white left wing.

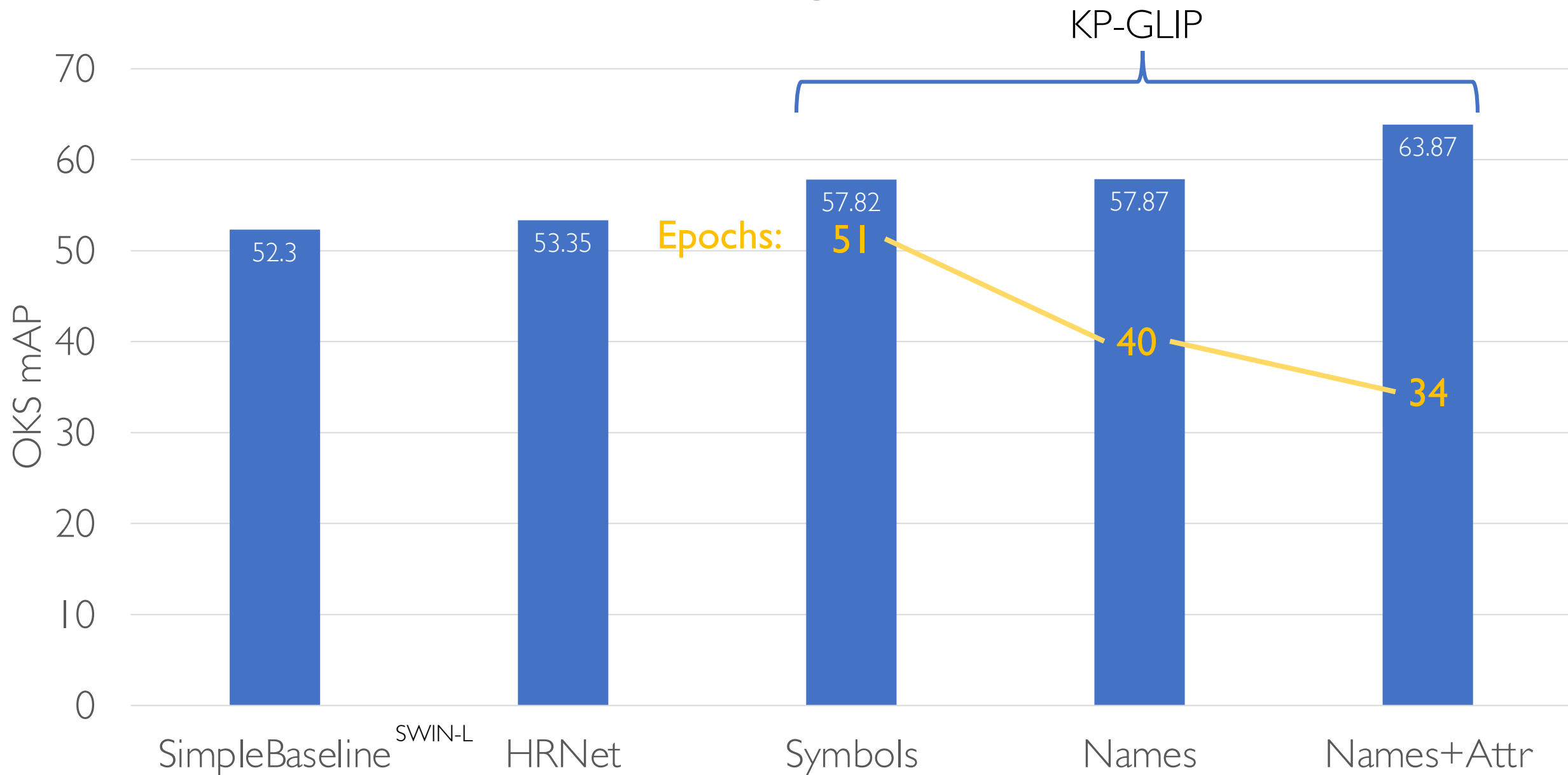
Finetuned N-Shot Learning



Finetuning Results



Finetuning Results



Conclusions

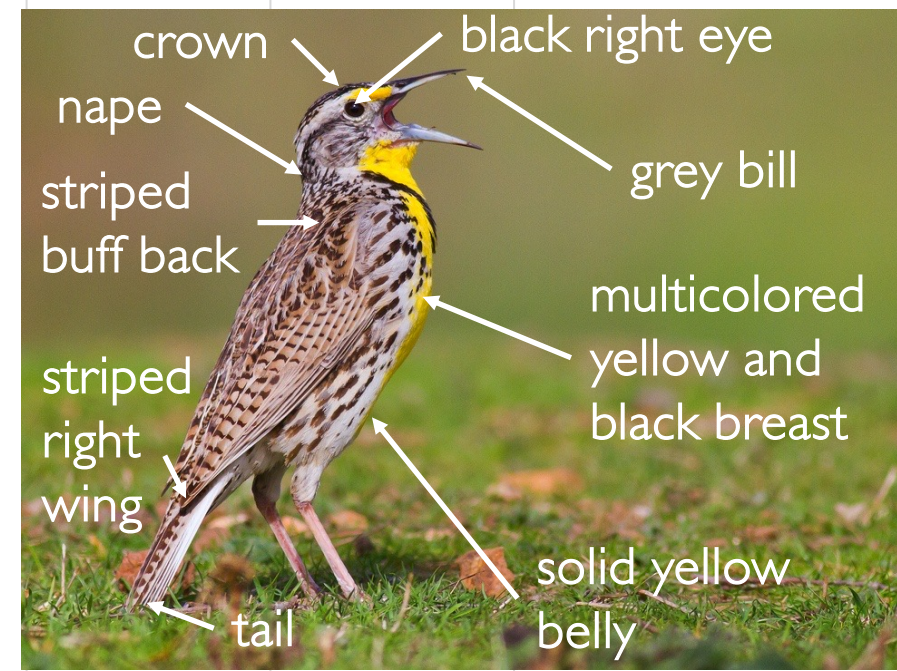
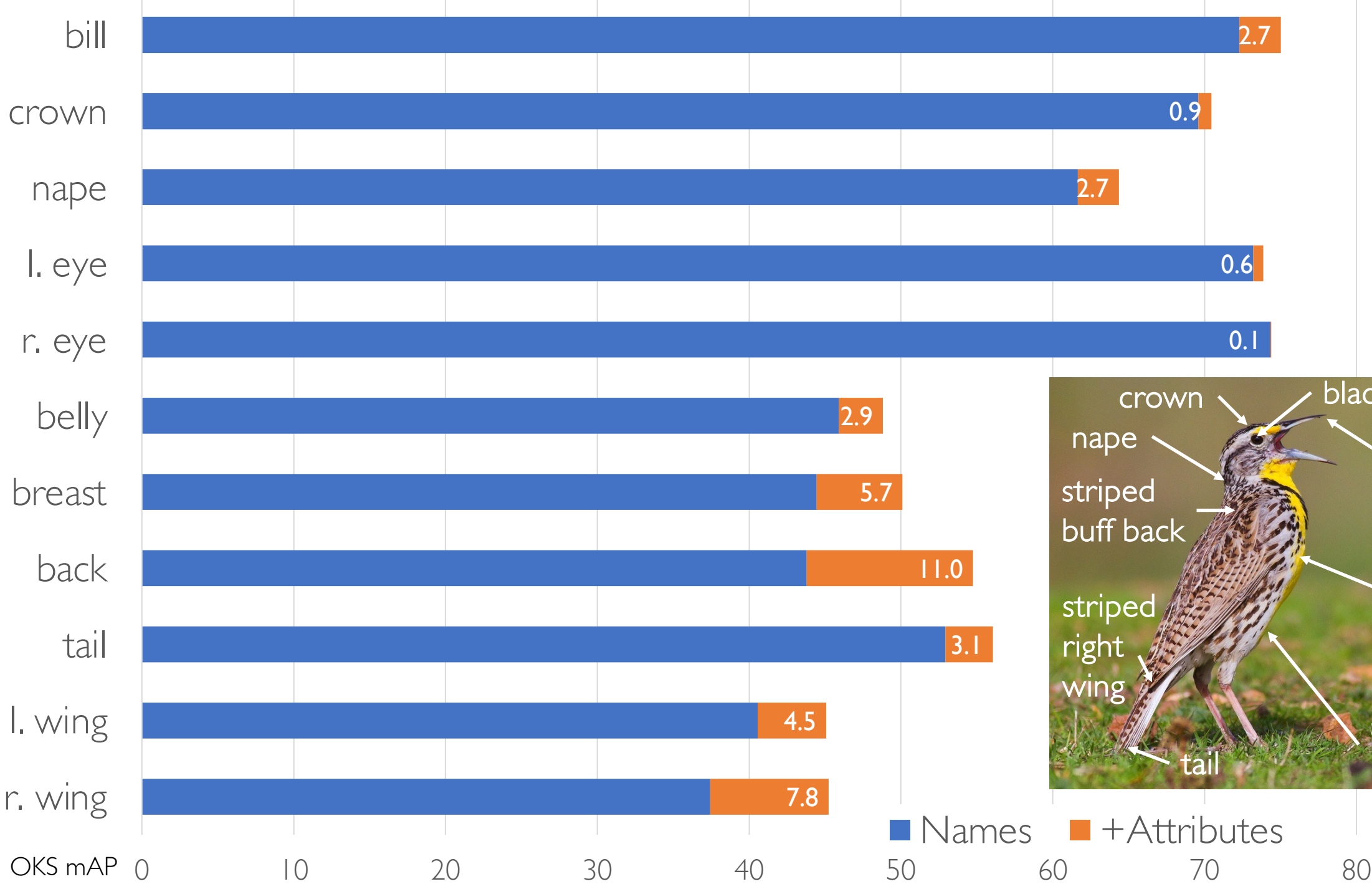
- Vision+Language models can excel at object keypoint detection
- Descriptive attributes leverage language and improve results

Conclusions

- Vision+Language models can excel at object keypoint detection
- Descriptive attributes leverage language and improve results

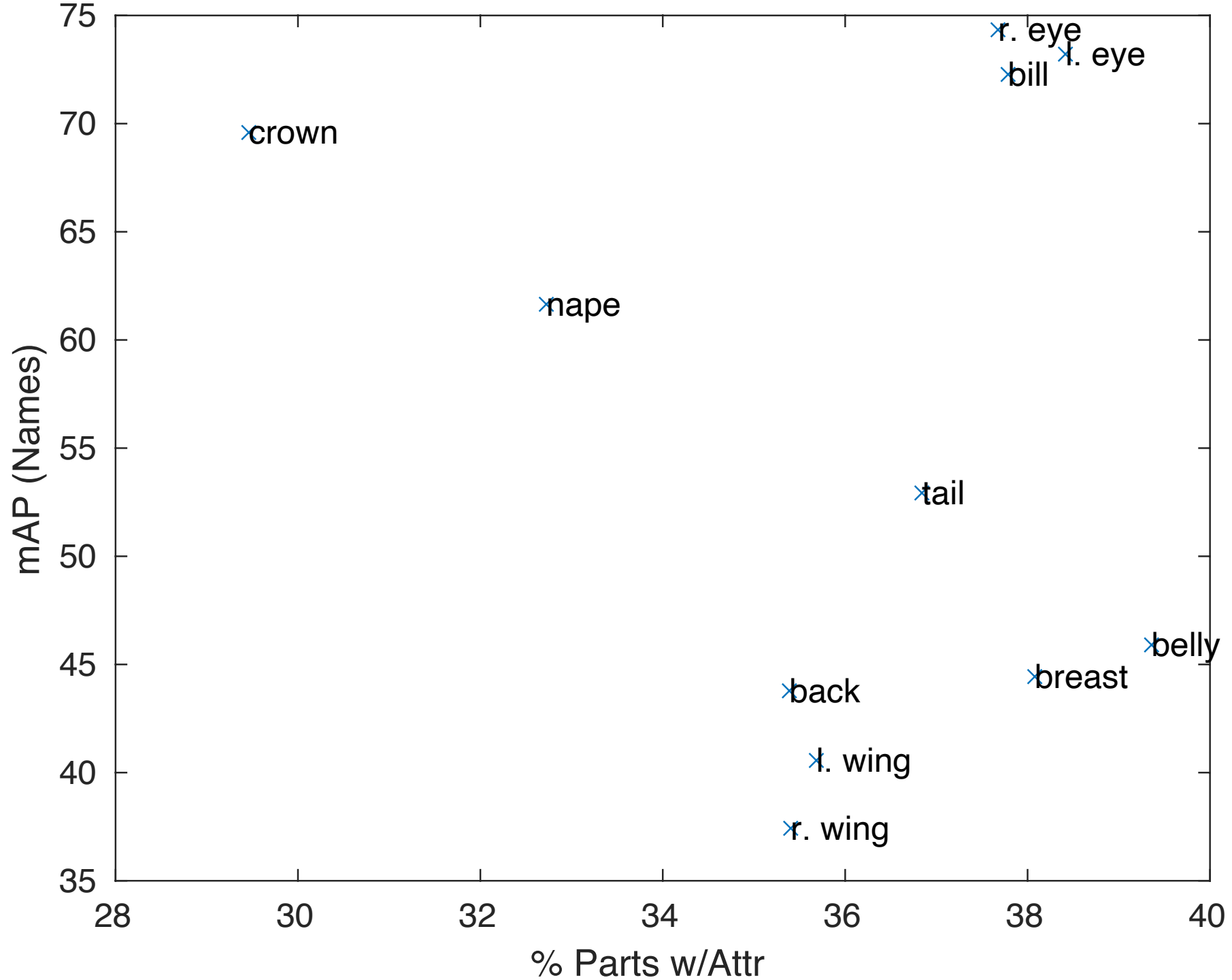
Future Work

- Jointly train with object detection
- Incorporate fine-grained species classification (“*barn swallow*” vs. “*tree swallow*” vs. “*cliff swallow*” vs. “*violet-green swallow*” vs. “*northern rough-winged swallow*”)



[NABirds/Davor Desancic]

■ Names ■ +Attributes



Validation Data

