Neglected Free Lunch
Learning Image Classifiers Using Annotation Byproducts

Dongyoon Han* Naver AI Lab
Junsuk Choe* Sogang University
Seonghyeok Chun DANTE COMPANY
John Joon Young Chung University of Michigan
Minsuk Chang Google
Sangdoo Yun
Jean Y. Song DGIST
Seong Joon Oh STAI University of Tübingen

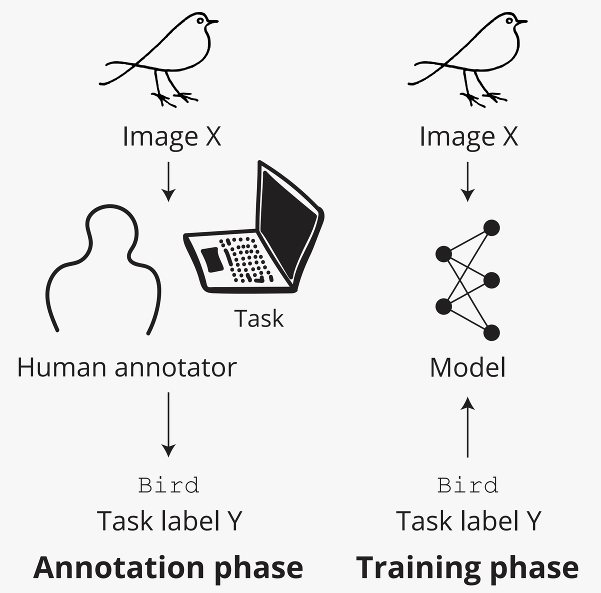
*Equal contribution

Funded by Naver and DGIST

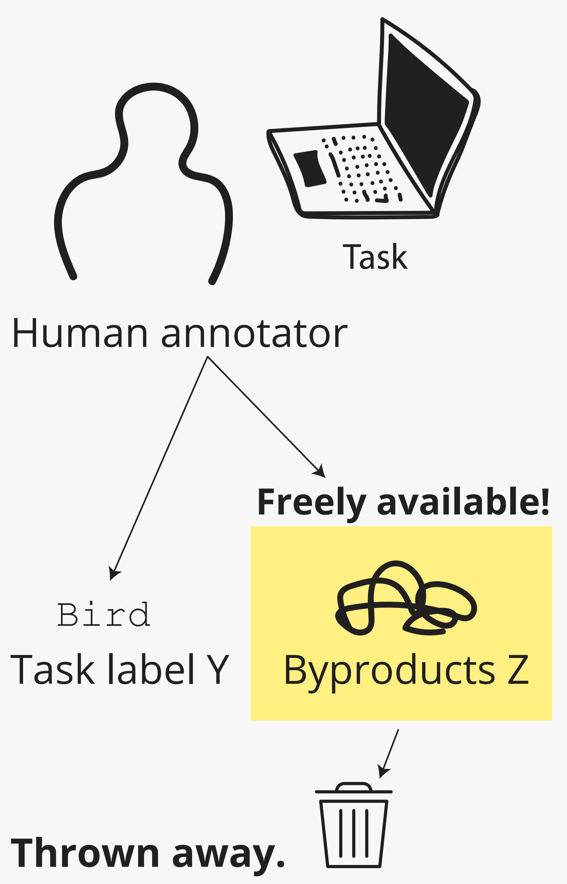
Motivation

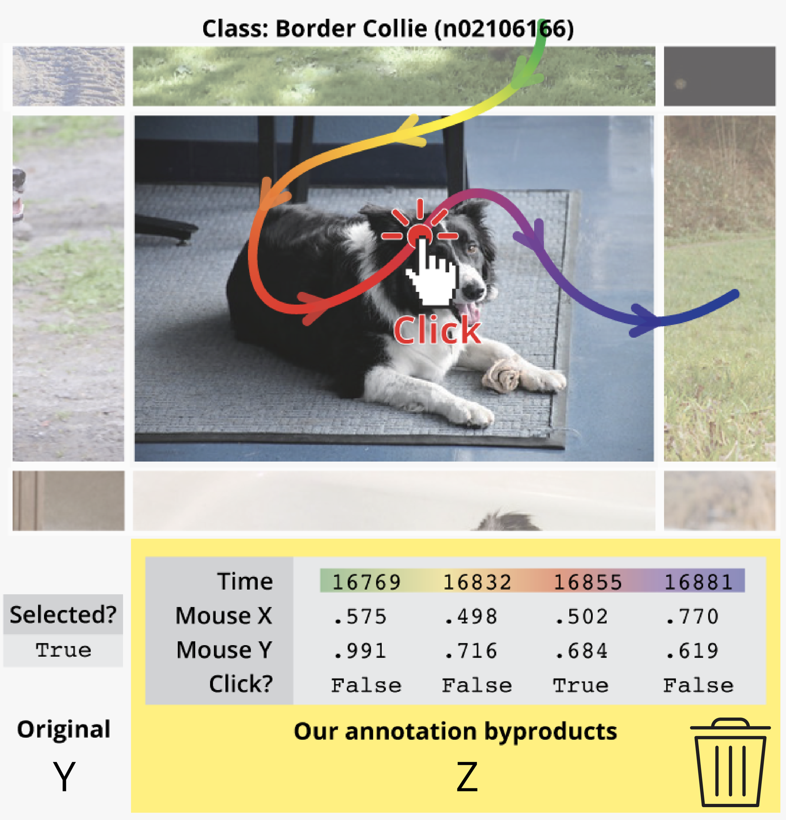
Supervised learning

Widely-used recipe:
(1) Collect Y for each X.
(2) Supervise model f with (X,Y).



Neglected bit: Annotation byproducts (AB)





Example ABs for ImageNet annotation.

Human-computer interactions generate traces.
Task label Y is only one of them.

Example byproducts Z:		Information they contain:
- Click locations - Mouse trajectory	→	Weak object location?
- Time to click - Correction history	→	Sample difficulty?
- Annotator ID - Task ID	→	Annotation bias?

Do ABs further improve models?

Main Message

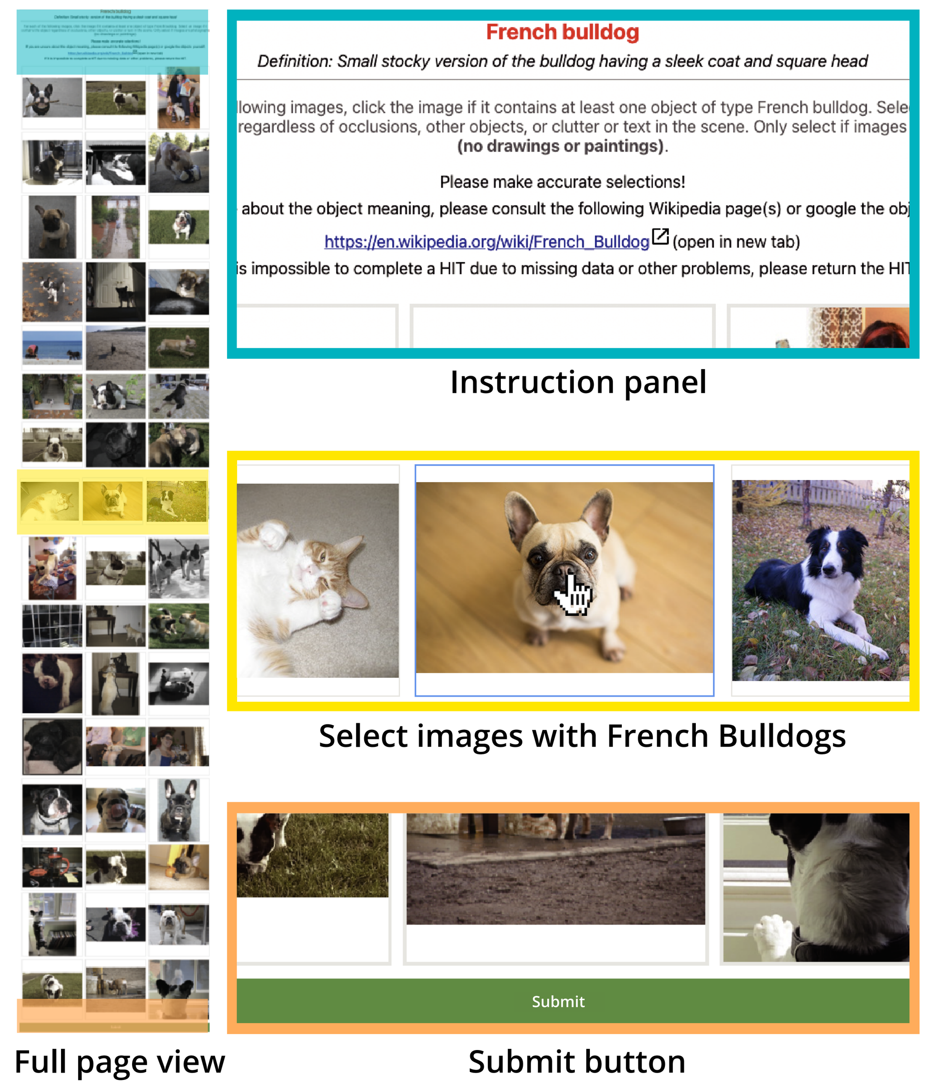
Are you *collecting annotations* for supervised learning?
You should definitely log *annotation byproducts*.
They may improve model performances *for free*.

ImageNet-AB Dataset: huggingface.co/datasets/coallaoh/ImageNet-AB
COCO-AB Dataset: huggingface.co/datasets/coallaoh/COCO-AB
Code: github.com/naver-ai/NeglectedFreeLunch

Collecting Annotation Byproducts

ImageNet-AB

1,272,225 / 1,281,167 (99.3%) ImageNet training images now have AB. 24,365 USD spent.

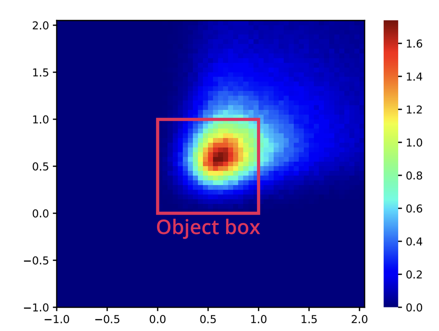


Replication of ImageNet annotation interface.


```
"imageID": "n01440764/n01440764_105",
"originalImageHeight": 375,
"originalImageWidth": 500,
"selected": true,                          Original Annotation
"imageHeight": 243,
"imageWidth": 243,
"imagePosition": {"x": 857, "y": 1976},
"hoveredRecord": [
  {"action": "enter", "time": 1641425051},
  {"action": "leave", "time": 1641425319}
],
"selectedRecord": [
  {"x": 0.540, "y": 0.473, "time": 1641425052}
],
"mouseTracking": [
  {"x": 0.003, "y": 0.629, "time": 1641425051},
  {"x": 0.441, "y": 0.600, "time": 1641425052}
],
"worker_id": "47DBDD543E",
"assignment_id": "3AMYWKA6YLE80HK9QYYHI2YEL2YO6L",
"page_idx": 3                              Annotation Byproducts
```


Use the last click location as a proxy of object location.

Example click locations.



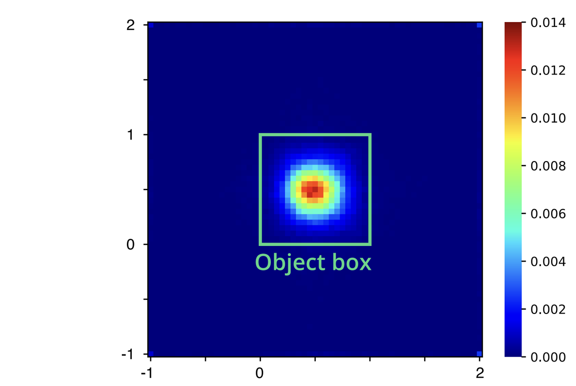
Distribution of clicks wrt GT boxes.

COCO-AB

82,765 / 82,783 (99.98%) COCO2014 training images now have AB. 9,936 USD spent.

Replication of COCO annotation interface.

Use the last icon location as a proxy of object location.

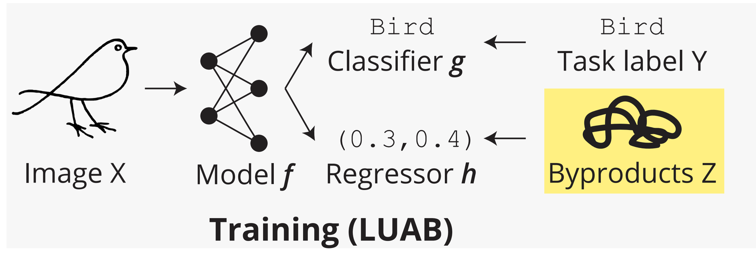


Distribution of icons wrt GT boxes.

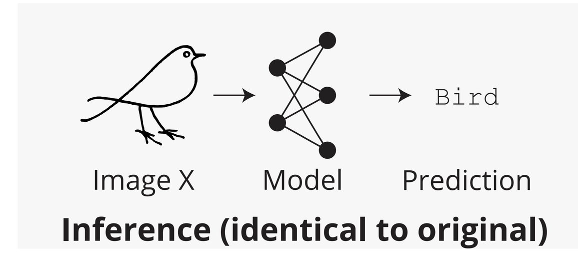
Learning Using AB (LUAB)

Special case of *Learning Using Privileged Information (LUPI)*.
This work: Focus on AB approximating object locations.

$$\min_{f,g,h} \mathcal{L}(g(f(X)),Y) + \lambda \| h(f(X)) - Z \|_{s1}$$



Training (LUAB)



Inference (identical to original)

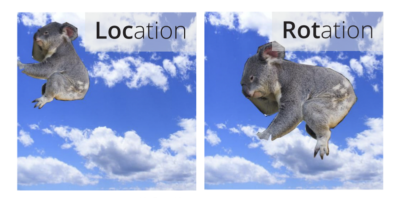
Results

ImageNet-AB + LUAB

Model	Params	IN-1K↑	IN-V2↑	IN-Real↑	IN-A↑	IN-C↑	IN-O↑	Sketch↑	IN-R↑	Cocc↑	ObjNet↑	SI-size↑	SI-loc↑	SI-rot↑	BGC-gap↓	BGC-acc↑
R18	11.7M	72.1	59.9	**79.6**	**2.0**	37.4	52.7	**22.0**	34.0	41.9	21.7	46.4	22.9	32.1	9.0	**22.1**
+LUAB	11.7M	**72.2**	**59.9**	79.6	1.9	**37.6**	**53.0**	21.6	**34.3**	**44.7**	**21.9**	**47.8**	**23.1**	**32.7**	**8.6**	20.4
R50	25.6M	77.4	65.2	83.5	**5.5**	43.8	56.7	25.4	37.8	53.7	27.8	53.9	31.9	40.1	6.3	26.7
+LUAB	25.6M	**77.5**	**65.2**	**83.8**	5.1	**44.7**	**57.0**	**25.7**	**38.2**	**55.1**	**28.5**	**55.6**	**33.5**	**40.9**	**5.6**	**27.4**
R101	44.5M	78.2	66.0	84.1	7.6	47.0	**60.7**	26.5	38.2	55.8	29.4	53.4	33.1	38.9	5.6	**30.2**
+LUAB	44.5M	**78.6**	**66.4**	**84.3**	**7.8**	**47.9**	60.5	**27.0**	**39.0**	**58.5**	**30.0**	**54.4**	**33.3**	**39.8**	**5.5**	28.2
R152	60.2M	79.0	67.2	84.5	9.5	49.5	62.0	27.6	39.6	58.8	30.5	53.9	33.3	38.6	6.6	27.2
+LUAB	60.2M	**79.2**	**67.2**	**84.8**	**9.5**	**49.9**	**62.1**	**27.6**	**39.7**	**59.0**	**31.3**	**55.5**	**34.2**	**40.6**	**5.8**	**31.6**
ViT-Ti	5.7M	72.8	60.7	80.7	7.9	**48.5**	52.3	20.5	32.8	63.8	23.1	46.3	23.8	33.9	8.2	13.9
+LUAB	5.7M	**72.9**	**60.8**	**80.9**	**8.4**	48.4	**52.9**	**21.1**	**33.8**	**64.2**	**23.7**	**47.4**	**25.4**	**34.7**	**7.8**	**14.4**
ViT-S	22.1M	80.3	69.1	86.0	20.0	60.3	53.4	29.4	42.3	73.8	31.2	54.5	32.0	39.5	6.4	17.4
+LUAB	22.1M	**80.6**	**69.7**	**86.4**	**22.8**	**61.2**	**55.1**	**30.0**	**43.0**	**74.1**	**32.3**	**55.1**	**33.7**	**39.6**	**5.9**	**18.7**
ViT-B	86.6M	81.6	70.3	86.6	26.1	64.1	58.0	33.0	45.7	76.0	31.7	56.6	35.1	41.3	6.4	18.1
+LUAB	86.6M	**82.5**	**71.9**	**87.4**	**31.1**	**66.0**	**58.5**	**35.5**	**48.4**	**77.5**	**35.0**	**57.1**	**36.8**	**41.6**	**5.6**	**23.9**

ID generalisation (IN-1K, IN-V2, IN-Real) · OOD generalisation (IN-A to ObjNet) · Spurious BG dependence (SI-size to BGC-acc)

LUAB improves ID & OOD gen. and reduces BG dependence.



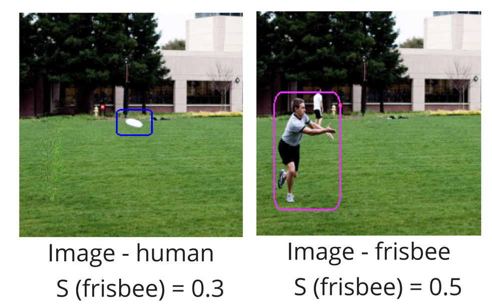
SI Scores

BG Challenge

COCO-AB + LUAB

Model	R18	Rand	LUAB	R50	Rand	LUAB	R152	Rand	LUAB
mAP↑	67.9	67.8	**68.0**	73.0	73.6	**74.2**	73.3	74.6	**75.4**
$V^{\min}$↓	51.8	52.1	**51.6**	47.6	47.3	**47.0**	47.4	47.8	**47.1**
V^{avg}↓	28.7	28.7	**28.4**	25.0	24.9	**24.5**	24.8	25.5	**24.7**

Model	ViT-Ti	Rand	LUAB	ViT-S	Rand	LUAB	ViT-B	Rand	LUAB
mAP↑	72.6	72.2	**72.7**	76.2	76.9	**77.3**	76.4	74.5	**77.5**
$V^{\min}$↓	49.1	48.9	**48.4**	47.1	46.9	**45.8**	46.6	47.1	**45.6**
V^{avg}↓	27.0	26.9	**26.8**	25.7	25.6	**24.6**	25.0	25.1	**24.5**



→ Vmin & Vavg ↑

- LUAB helps ID generalisation.
- LUAB reduces spurious BG dependence.